When is TSLS *Actually* LATE?[*]

Christine Blandhol[†] John Bonney[‡]

Magne Mogstad[§] Alexander Torgovitsky[¶]

August 8, 2022

Abstract

Linear instrumental variable estimators, such as two-stage least squares (TSLS), are commonly interpreted as estimating positively weighted averages of causal effects, referred to as local average treatment effects (LATEs). We examine whether the LATE interpretation actually applies to the types of TSLS specifications that are used in practice. We show that if the specification includes covariates—which most empirical work does—then the LATE interpretation does not apply in general. Instead, the TSLS estimator will, in general, reflect treatment effects for both compliers *and* always/never-takers, and some of the treatment effects for the always/never-takers will *necessarily* be negatively weighted. We show that the only specifications that have a LATE interpretation are "saturated" specifications that control for covariates nonparametrically, implying that such specifications are both sufficient and *necessary* for TSLS to have a LATE interpretation, at least without additional parametric assumptions. This result is concerning because, as we document, empirical researchers almost never control for covariates nonparametrically, and rarely discuss or justify parametric specifications of covariates. We develop a decomposition that quantifies the extent to which the usual LATE interpretation fails. We apply the decomposition to four empirical analyses and find strong evidence that the LATE interpretation of TSLS is far from accurate for the types of specifications actually used in practice.

[*] We thank Richard Boylan, Deniz Dutz, Mikkel Gandil, Tymon Słoczyński, and Chris Walters for helpful comments.

[†] Department of Economics, Princeton University.

[‡] Department of Economics, Stanford University. Research supported by National Science Foundation graduate research fellowship.

[§] Kenneth C. Griffin Department of Economics, University of Chicago; Statistics Norway; NBER.

[¶] Kenneth C. Griffin Department of Economics, University of Chicago. Research supported by National Science Foundation grant SES-1846832.

1 Introduction

Instrumental variable (IV) strategies are widely used for causal inference in economics, political science, sociology, epidemiology, and other fields. Since the work of Imbens and Angrist (1994), it has been increasingly common to interpret linear IV estimators as estimating a local average treatment effect (LATE), or at least a positively weighted average of LATEs. The scientific value of this interpretation has been extensively analyzed and debated (e.g. Robins and Greenland, 1996; Heckman, 1997; Angrist and Imbens, 1999; Deaton, 2010; Imbens, 2010; Swanson and Hernán, 2014).

In this paper we focus on a more practical question that is both distinct from, and primary to, the value of the LATE interpretation: does the LATE interpretation even apply to the types of specifications that empirical researchers use in practice? In Section 2, we show that if the IV specification includes covariates, then the answer is, in general, "no." The linear IV estimator with covariates is composed of treatment effects for both compliers *and* always-takers, and some always-taker treatment effects are always negatively weighted.

Our finding challenges the claim by Angrist and Pischke (2009, pg. 173) that

> *2SLS with covariates produces an average of covariate-specific LATEs... These results provide a simple casual* [typo in original] *interpretation for 2SLS in most empirically relevant settings.*

Their assertion is based on a "saturated" two stage least squares (TSLS) specification that controls for covariates nonparametrically, described by Angrist and Pischke (2009) as the "saturate and weight approach" (Theorem 4.5.1; originally Theorem 3 in Angrist and Imbens, 1995). Our results show that this type of saturated specification is not only sufficient for TSLS with covariates to be interpretable as an average of covariate-specific LATEs, it is also *necessary*, at least without additional parametric assumptions.[1]

In Section 2, we report the results of a survey on the specification of linear IV estimators in published empirical papers in economics. Of the 99 papers in our survey that use a linear IV estimator with covariates, we find only a single paper (Chamberlain and Imbens, 2004) that used a saturated specification. The implication for the 98 other papers is that they may not be estimating an average of covariate-specific LATEs. In fact, they may be estimating a quantity that doesn't even satisfy the minimal requirement of being a positively weighted average of subgroup-specific treatment effects, a

[1] As we show in Section 4.2, this statement remains true even if treatment effects are constant across both unobservable groups and observed covariates. This contrasts with recent results on two-way fixed effects models (e.g. Goodman-Bacon, 2021; Sun and Abraham, 2021), which point out interpretation problems that arise in event studies if there are heterogeneous treatment effects due to observables (in particular, cohorts).

property we describe as *weakly causal.*

In Section 3, we provide a formal definition of a weakly causal estimand and develop a general sufficient and necessary characterization of estimands that are weakly causal. The characterization has two components. First, a weakly causal estimand cannot depend on the levels of potential outcomes while holding treatment effects (differences) constant, a phenomenon we describe as *level-dependence.* Second, a weakly causal estimand should not apply negative weight to the treatment effects for any subgroup. In Section 4, we specialize these results to a large class of TSLS estimands.

We start by considering the simpler case with treatment effects that are linear in the treatment intensity and constant across both unobservable and observable groups. For this case we show that the single sufficient and necessary condition for the TSLS estimand to be weakly causal is that the TSLS specification has *rich covariates*, in the sense that it exactly reproduces the conditional mean of the instrument.[2] Specifications that are saturated in covariates, such as the Angrist and Pischke (2009) "saturate and weight" specification, will always have rich covariates. But a non-saturated specification only has rich covariates if an implicit parametric functional form assumption happens to be correct. If it is not correct, our results show that the resulting TSLS estimand will necessarily depend on potential outcome levels (level-dependence), and so will not be weakly causal. In the constant effects case, rich covariates can be replaced by traditional linearity assumptions on mean potential outcomes (e.g. Heckman and Robb, 1985).

Next, we allow for heterogeneous, nonlinear treatment effects and impose the Imbens and Angrist (1994) monotonicity condition. A rich covariate specification is still necessary for the TSLS estimand to be weakly causal, but it is no longer sufficient. Given rich covariates, we show that the additional sufficient and necessary condition is that the first stage specification is *monotonicity-correct*, meaning that it correctly reproduces the direction of the monotonicity assumption conditional on any value of the covariates. When the first stage is not monotonicity-correct, the TSLS estimand will always reflect the contribution of a negatively weighted subgroup, and so will not be weakly causal. In the heterogeneous treatment effects case, rich covariates *cannot* be replaced by traditional linearity assumptions on mean potential outcomes.

Our sufficient conditions extend those by Angrist and Imbens (1995, Theorem 3), Abadie (2003, Propositions 5.1 and 5.2), Kolesár (2013), and Słoczyński (2022). More importantly, unlike these authors, we show that rich covariates and a monotonicity-

[2] In this regard, our results differ markedly from those of Słoczyński (2022), who *assumes* rich covariates and examines negative weight problems that can only arise under heterogeneous treatment effects due to non-monotonicity.

correct first stage are also *necessary* for a TSLS estimand to be weakly causal. The implication is that the Angrist and Pischke (2009) interpretation of TSLS as a positively weighted average of LATEs is fragile. Unless one happens to specify the relationship between the instruments and covariates correctly, and unless one happens to include enough interactions between the instrument and covariates in the first stage, the TSLS estimand will not be a positively weighted average of LATEs. Assuming constant treatment effects allows one to omit first stage interactions, but still does not mean TSLS will have a causal interpretation without correctly specified functional forms.

Although our survey turned up only a single paper that used a TSLS specification with saturated covariates, we found many papers that nevertheless invoked the widespread LATE interpretation. Our results draw this interpretation into question. Yet the magnitude by which the interpretation fails in practice is ultimately an empirical question.

In Section 5, we develop a decomposition that quantifies the extent to which a TSLS estimand fails to be weakly causal. The decomposition contains three terms. The first is a level-dependence term that reflects how not having rich covariates makes the TSLS estimand depend on the levels of the potential outcomes, rather than solely on treatment effects. The second term consists of negatively weighted treatment effects that are created by a monotonicity-incorrect first stage, even with rich covariates. The third term consists of positively weighted treatment effects. The TSLS estimand is weakly causal if and only if the first two terms in the decomposition are zero. All three terms are identified and can be estimated, enabling quantification of the extent to which a given TSLS estimand fails to be weakly causal.

In Section 6, we apply the decomposition to four IV analyses: Gelbach (2002), Dube and Harish (2020), Card (1995), and Angrist and Krueger (1991). In the first two applications, we find strong evidence that covariates are not rich, leading to TSLS estimates that aren't weakly causal even under a constant, linear treatment effects assumption, at least not without additional parametric assumptions. In the third and fourth applications, we find that the covariates are rich enough to prevent severe level-dependence, but that the first stages are not monotonicity-correct, leading the TSLS estimand to reflect negatively weighted treatment effects. Our findings in all four applications suggest that the LATE interpretation of TSLS is far from accurate for the types of specifications actually used in practice.

In the concluding section, we summarize our key findings and discuss the implications for empirical research. Taken together, our findings show that for TSLS to meet even the weakly causal criterion requires either making parametric assumptions or controlling for covariates nonparametrically. The implication is that TSLS does not possess a privileged causal interpretation compared to more principled methods for estimating the LATE that have been proposed in the literature. These include fully parametric methods (Imbens and Rubin, 1997; Hirano et al., 2000; Yau and Little, 2001; Słoczyński et al., 2022), semiparametric methods (Abadie, 2003; Tan, 2006, 2010; Hong and Nekipelov, 2010), and nonparametric methods (Frölich, 2007; Ogburn et al., 2015; Heiler, 2021; Sun and Tan, 2021), including machine learning based approaches (Chernozhukov et al., 2018; Athey et al., 2019; Singh and Sun, 2022). Our survey of the empirical literature shows that these methods are not widely used. Our findings on TSLS suggest that maybe they should be.

2 Overview and implications for empirical practice

In this section we demonstrate our main points in the special case of a binary treatment and binary instrument with heterogeneous treatment effects.

2.1 IV with covariates is not LATE

Let $T \in \{0, 1\}$ be a binary treatment and $Z \in \{0, 1\}$ be a binary instrument. The outcome is Y with potential outcomes $Y(0)$ and $Y(1)$ related via $Y = (1 - T)Y(0) + TY(1)$. Potential treatment states are $T(0)$ and $T(1)$ with $T = (1 - Z)T(0) + ZT(1)$. The vector of covariates is X.

Assume that Z is conditionally exogenous in the sense of being independent of $(Y(0), Y(1), T(0), T(1))$ conditional on X. Suppose that the Imbens and Angrist (1994) monotonicity conditions holds so that $\mathbb{P}[T(1) \geq T(0)] = 1$.[3] The monotonicity condition implies that the group variable $G \equiv (T(0), T(1))$ can take three values with non-zero probability: $G = (0, 0) \equiv \textsc{nt}$ are the never-takers, $G = (0, 1) \equiv \textsc{cp}$ are the compliers, and $G = (1, 1) \equiv \textsc{at}$ are the always-takers.

Consider a linear IV regression with outcome variable Y, endogenous variable T, excluded instrument Z, and a vector of control variables X that includes a constant. The IV estimand (the population coefficient on T) is given by

$$\beta_{\text{iv}} = \frac{\mathbb{E}[Y\tilde{Z}]}{\mathbb{E}[T\tilde{Z}]}, \quad \text{where} \quad \tilde{Z} \equiv Z - \mathbb{L}[Z|X] \tag{1}$$

are the residuals from a regression of Z on X, and

$$\mathbb{L}[Z|X] \equiv X' \, \mathbb{E}[XX']^{-1} \, \mathbb{E}[XZ]$$

[3] This is the strongest form of the monotonicity condition one can contemplate here. Weaker forms that allow the ordering of $T(0)$ and $T(1)$ to vary with X are considered in Section 4.4.

are the population fitted values from regressing (linearly projecting) Z onto X.[4] The IV estimand β_{iv} is often interpreted as reflecting the average treatment effect among complier groups with weights that vary by the probability of compliance given covariates. The following proposition shows that this is not true in general.

Proposition 1. Suppose that $\mathbb{E}[Y(t)|X] = \eta_t' X$ for some (unknown) parameters η_t, $t = 0, 1$.[5] Let $\Delta(\textsc{cp}, x) \equiv \mathbb{E}[Y(1) - Y(0)|G = \textsc{cp}, X = x]$ and $\Delta(\textsc{at}, x) \equiv \mathbb{E}[Y(1) - Y(0)|G = \textsc{at}, X = x]$ denote the conditional average treatment effects for the compliers and always-takers, respectively. Then

$$\begin{aligned} \beta_{\text{iv}} &= \mathbb{E}[\omega(\textsc{cp}, X)\Delta(\textsc{cp}, X)] + \mathbb{E}[\omega(\textsc{at}, X)\Delta(\textsc{at}, X)], \qquad (2) \\ \text{where} \quad \omega(\textsc{cp}, X) &\equiv \mathbb{E}[Z|X]\,(1 - \mathbb{L}[Z|X])\,\mathbb{P}[G = \textsc{cp}|X]\,\mathbb{E}[\tilde{Z}T]^{-1} \\ \text{and} \quad \omega(\textsc{at}, X) &\equiv \mathbb{E}[\tilde{Z}|X]\,\mathbb{P}[G = \textsc{at}|X]\,\mathbb{E}[\tilde{Z}T]^{-1}. \end{aligned}$$

If $\mathbb{E}[\tilde{Z}T] > 0$, then the complier weights $\omega(\textsc{cp}, X)$ are negative if and only if $\mathbb{L}[Z|X] > 1$. The always-taker weights $\omega(\textsc{at}, X)$ are strictly negative with positive probability unless $\mathbb{E}[\tilde{Z}|X] = 0$ deterministically.

Proposition 1 shows that in general β_{iv} reflects not only the compliers, but also the always-takers. If $\mathbb{E}[T\tilde{Z}] > 0$, so that the first stage coefficient is positive, then the weights on the always-takers have the same sign as the random variable $\mathbb{E}[\tilde{Z}|X] = \mathbb{E}[Z|X] - \mathbb{L}[Z|X]$. Because $\mathbb{L}[Z|X]$ is the best linear approximation to $\mathbb{E}[Z|X]$, the difference $\mathbb{E}[\tilde{Z}|X]$ always takes negative values whenever it is not deterministically zero. Thus, whenever $\mathbb{L}[Z|X] \neq \mathbb{E}[Z|X]$, the IV estimand incorporates negatively weighted treatment effects for some groups, which means that it fails to satisfy even a minimal condition for "being causal."[6]

In order for the LATE interpretation to hold, it is necessary that $\mathbb{L}[Z|X] = \mathbb{E}[Z|X]$, a condition we call *rich covariates*. Specifications that are saturated in covariates, such as "saturate and weight" (Angrist and Pischke, 2009), have rich covariates. If Z and X are independent, as can be the case in some controlled and natural experiments, then any specification with a constant will have rich covariates. Outside of these two cases, having rich covariates is a parametric assumption. If it fails, then the IV estimand β_{iv} reflects not just compliers, but also negatively weighted always-takers.

[4] The proof of (1) is a special case of Proposition 6 ahead.

[5] This additional assumption is made in order to simplify the weights. Removing the assumption only *amplifies* the negative interpretation issues exposed by Proposition 1. Our general results in Section 4 do not maintain this assumption.

[6] It's also possible that $\mathbb{E}[T\tilde{Z}] < 0$, so that the first stage coefficient has the opposite sign of that suggested by the monotonicity condition (see Section 4.3). This does not change the conclusion that there will always be some negatively weighted always-takers unless $\mathbb{E}[\tilde{Z}|X]$ is deterministically zero.

There is no reason to expect, a priori, that the weights on the always-taker treatment effects in (2) will be small in magnitude. In many applications, the proportion of always-takers, $\mathbb{P}[G = \textsc{at}|X]$, can be expected to be considerably larger than the proportion of compliers, $\mathbb{P}[G = \textsc{cp}|X]$. As a consequence, even negative values of $\mathbb{E}[\tilde{Z}|X]$ that are small in magnitude can produce large negative weights on the always-taker treatment effects. These magnitudes are ultimately an empirical matter. The decomposition we develop in Section 5, and apply in Section 6, provides a way to quantify the impact that a failure of rich covariates has on the causal interpretation of β_{iv}.

Decomposition (2) is not the only possible decomposition. Instead of interpreting β_{iv} as a weighted average of compliers and always-takers, one can interpret it as a weighted average of compliers and never-takers, or of all three groups, as shown in the next proposition.

Proposition 2. Suppose that $\mathbb{E}[Y(t)|X] = \eta_t' X$ for some (unknown) parameters η_t, $t = 0, 1$. Let $\Delta(\textsc{nt}, x) \equiv \mathbb{E}[Y(1) - Y(0)|G = \textsc{nt}, X = x]$ denote the conditional average treatment effect for the never-takers. Then for any real number ϵ,

$$\begin{aligned} \beta_{\text{iv}} &= \mathbb{E}[\omega_\epsilon(\textsc{cp}, X)\Delta(\textsc{cp}, X)] + \mathbb{E}[\omega_\epsilon(\textsc{at}, X)\Delta(\textsc{at}, X)] + \mathbb{E}[\omega_\epsilon(\textsc{nt}, X)\Delta(\textsc{nt}, X)], \\ \text{where} \quad & \omega_\epsilon(\textsc{cp}, X) \equiv \Big(\epsilon\, \mathbb{E}[\tilde{Z}|X] + \mathbb{L}[Z|X](1 - \mathbb{E}[Z|X])\Big)\, \mathbb{P}[G = \textsc{cp}|X]\, \mathbb{E}[\tilde{Z}T]^{-1}, \\ & \omega_\epsilon(\textsc{at}, X) \equiv \epsilon\, \mathbb{E}[\tilde{Z}|X]\, \mathbb{P}[G = \textsc{at}|X]\, \mathbb{E}[\tilde{Z}T]^{-1}, \\ \text{and} \quad & \omega_\epsilon(\textsc{nt}, X) \equiv (\epsilon - 1)\, \mathbb{E}[\tilde{Z}|X]\, \mathbb{P}[G = \textsc{nt}|X]\, \mathbb{E}[\tilde{Z}T]^{-1}. \end{aligned}$$

Each choice of ϵ in Proposition 2 provides a different interpretation of β_{iv}, with Proposition 1 corresponding to $\epsilon = 1$. However, unless $\mathbb{E}[\tilde{Z}|X] = \mathbb{E}[Z|X] - \mathbb{L}[Z|X] = 0$, any such choice involves either the always-takers or the never-takers, or both, and applies negative weights to both groups for some values of X, as well as potentially negative weights to the compliers. Only in specifications with rich covariates is β_{iv} a positively weighted average among compliers alone.

2.2 Intuition

The intuition behind Propositions 1 and 2 can be seen by writing the numerator of β_{iv} as

$$\mathbb{E}[Y\tilde{Z}] = \mathbb{E}\left[\mathbb{E}\left[Y\tilde{Z}|X\right]\right] = \mathbb{E}\left[\overbrace{\mathbb{C}[Y, Z|X]}^{\text{only contains complier treatment effects}}\right] + \mathbb{E}\left[\underbrace{\mathbb{E}[Y|X]}_{\text{contains all three groups}}\mathbb{E}[\tilde{Z}|X]\right]. \qquad (3)$$

The first term in (3) is the average of the numerator of a nonparametric IV specification that *conditions* on X. The argument in Imbens and Angrist (1994) shows that this term is equal to an average of scaled LATEs, which only reflects treatment effects for the compliers. It is the second term of (3) that causes problems. This term reflects the difference between nonparametric conditioning and linear projection.

When covariates are not rich, so that $\mathbb{E}[\tilde{Z}|X] \neq 0$, the second term in (3) generally depends on $\mathbb{E}[Y|X]$, a quantity which is determined not only by compliers, but also by always-takers and never-takers. This creates "level-dependence" in β_{iv} because the always-takers always have $Y = Y(1)$ and the never-takers always have $Y = Y(0)$. Thus, β_{iv} depends on the *levels* of the always-taker and never-taker potential outcomes, rather than the *difference*, $Y(1) - Y(0)$. As we show in Section 3, level-dependent estimands do not have a causal interpretation because the levels can lead β_{iv} to have the "wrong sign."

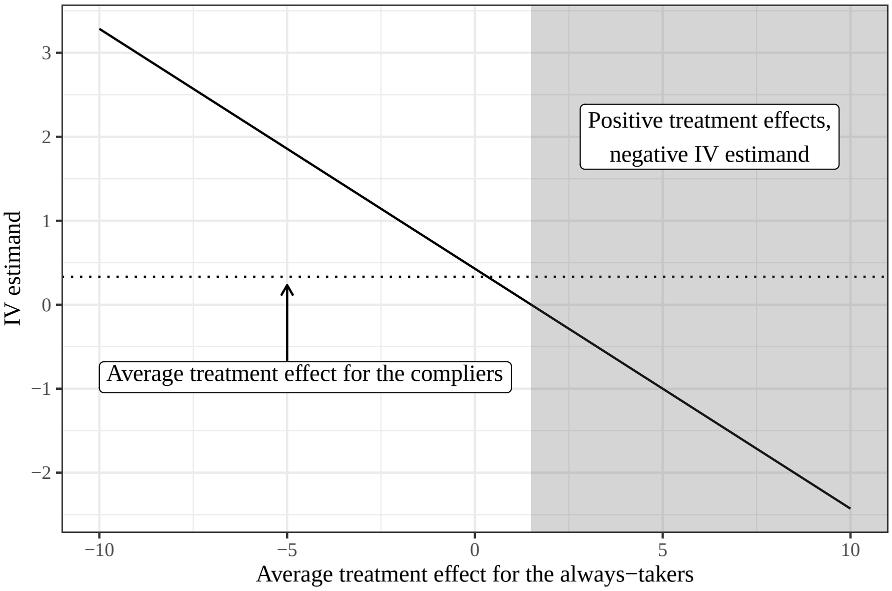
The expression in Proposition 1 arises from centering the term $\mathbb{E}[Y|X]$ in (3) around $\mathbb{E}[Y(0)|X]$. The simplifying linearity assumption implies that $\mathbb{E}[Y(0)|X] = \eta_0' X$ is uncorrelated with $\mathbb{E}[\tilde{Z}|X]$. Thus, since never-takers always have $Y = Y(0)$, the centering removes the average untreated outcome for the never-takers, leaving only a weighted average of the complier and always-taker treatment effects. Alternatively, we can center around $\mathbb{E}[Y(1)|X] = \eta_1' X$, which leaves a weighted average of the complier and never-taker treatment effects. Both decompositions are equally valid ways to rewrite a single number (β_{iv}) as a weighted average of three other numbers ($\Delta(\text{CP}, X)$, $\Delta(\text{AT}, X)$, and $\Delta(\text{NT}, X)$). Taking an ϵ–weighted average of these two decompositions yields the expression in Proposition 2, which covers all possible decompositions in this specific case.

The theory we develop in Section 3 is designed to handle this type of non-uniqueness in decomposition and determine, in a general setting, necessary conditions for the existence of *some* "good" decomposition. For the simplified case considered here, with a binary treatment, a binary instrument, and the linearity assumption $\mathbb{E}[Y(d)|X] = \eta_d' X$, this type of analysis can be done directly, as in Proposition 2. Our analysis of more general TSLS specifications in Section 4 shows that the necessity of rich covariates for a causal interpretation is a conclusion that applies more broadly.

2.3 Numerical illustration

As a simple illustration of these results, suppose that $X \in \{(1, -1), (1, 0), (1, 1)\}$ with equal probability, where the first component corresponds to a constant. Then suppose

Figure 1: IV with covariates is not LATE



that

$$\mathbb{E}[Z|X=x] = \mathbb{P}[Z=1|X=(1,x)] = \begin{cases} 4/5 & \text{if } x \in \{-1,1\} \\ 2/5 & \text{if } x = 0 \end{cases}.$$

Regressing Z onto X yields the constant regression line:

$$\mathbb{L}[Z|X] = X' \, \mathbb{E}[XX']^{-1} \, \mathbb{E}[XZ] = 2/3,$$

so that $\mathbb{E}[\tilde{Z}|X] = \mathbb{E}[Z|X] - \mathbb{L}[Z|X] \neq 0$ and is both positive and negative with non-zero probability.

Suppose that the conditional group share probabilities are given by:

$$\begin{aligned} &(\text{never-takers}) && \mathbb{P}[G = \textsc{nt}|X=(1,x)] = 1/3 \\ &(\text{compliers}) && \mathbb{P}[G = \textsc{cp}|X=(1,x)] = 1/6 + |x|/6 \\ &(\text{always-takers}) && \mathbb{P}[G = \textsc{at}|X=(1,x)] = 1/2 - |x|/6. \end{aligned}$$

Simplifying the algebra in Proposition 1 yields

$$\omega(\textsc{cp},(1,x)) = \begin{cases} 12/7, & \text{if } |x| = 1 \\ 3/7, & \text{if } x = 0 \end{cases} \quad \text{and} \quad \omega(\textsc{at},(1,x)) = \begin{cases} 6/7, & \text{if } |x| = 1 \\ -18/7, & \text{if } x = 0 \end{cases}.$$

For simplicity, assume that $Y(0) = 0$, so that treatment effects are determined solely by $Y(1)$, and that $\mathbb{E}[Y(1)|G = \textsc{cp}, X = x] \equiv \mu(\textsc{cp})$ and $\mathbb{E}[Y(1)|G = \textsc{at}, X = x] = \mu(\textsc{at})$ do not depend on x. Then Proposition 1 shows that

$$\beta_{\text{iv}} = \frac{9}{7}\mu(\textsc{cp}) - \frac{2}{7}\mu(\textsc{at}).$$

Figure 1 shows the value of β_{iv} as a function of $\mu(\textsc{at})$, keeping $\mu(\textsc{cp}) = 1/3$. If it were true that LATE only reflects the compliers, then we would expect to see a flat line, so that the IV estimand doesn't depend on the treatment effect for the always-takers. Not only is the line not flat, it slopes down. This means that the IV estimand can be negative even when both the compliers and the always-takers have positive treatment effects.

2.4 Survey on IV specifications used in empirical work

Propositions 1 and 2 show that using an IV specification that is saturated in covariates is important for the LATE interpretation asserted by Angrist and Pischke (2009). To get a sense of how common it is to saturate in covariates, we surveyed the specifications used in the empirical economics literature.

Our sample was constructed by searching the Web of Science Database for articles published between January 2000 and October 2018 containing the words "instrument" or "instrumental variable" in the abstract, title, or topic words. We restricted the search to the following five journals: *Journal of Political Economy*, *American Economic Review*, *Quarterly Journal of Economics*, *Review of Economic Studies*, and *Econometrica*. In total, 266 articles matched our search criteria.

We restricted our attention to papers that use at least one IV specification in an empirical application. This produced 122 papers; the other 144 papers not included were either methodological papers without an empirical application, or were papers that used the word "instrument" in a different context, such as to describe a policy or financial instrument. Column (1) of Table 1 tabulates the papers used in our survey by the journal in which they were published.

Column (2) shows that over 92% of the papers in our survey use TSLS (including exactly identified linear IV) for at least some of their results. Column (3) counts the subset of the papers in column (2) for which *all* TSLS specifications in the main body of the paper include at least one covariate, or the authors explicitly state the exogeneity assumption for the instrument as conditional on covariates.[7] Comparing columns (2)

[7] Another possible justification for including covariates is to improve statistical precision. This motivation was rarely stated explicitly in the papers in our survey. While it is difficult to infer researchers' unstated

Table 1: IV papers by journal and type

	(1)	(2)	(3)	(4)
	All papers	Papers using TSLS	Papers using TSLS with covariates	Papers using TSLS with covariates, referring to LATE
American Economic Review	100% 44	95% 42	82% 36	27% 12
Quarterly Journal of Economics	100% 28	93% 26	86% 24	14% 4
Journal of Political Economy	100% 23	91% 21	83% 19	30% 7
Econometrica	100% 15	73% 11	73% 11	27% 4
Review of Economic Studies	100% 12	100% 12	75% 9	25% 3
All	100 % 122	92% 112	81% 99	25% 30

and (3) shows that using covariates in TSLS is extremely common practice; only 13 out of the 112 papers that use TSLS include any specifications without covariates. Column (4) shows that almost a third of the papers that use TSLS with covariates also explicitly use the phrases "compliers," "local average treatment effect," or "LATE" to describe their results.

In Table 2, we categorize the papers in column (3) of Table 1 by the TSLS specifications they use. Column (2) shows that only 5% of the papers use any specification that is saturated in covariates. These are typically preliminary specifications with only a set of fixed effects. Column (3) shows that every paper uses at least one specification that is *not* saturated in covariates, with only one exception. The one exception is Chamberlain and Imbens (2004). Column (4) shows that those authors also saturate the first stage in both the covariates and the instruments, as prescribed by Angrist and Pischke's (2009) "saturate and weight" specification.

2.5 Implications for empirical practice

Avoiding the conclusion of Propositions 1 and 2 requires choosing a specification with rich covariates, i.e. one that ensures $\mathbb{L}[Z|X] = \mathbb{E}[Z|X]$.

reasons for choosing particular specifications, it seems unlikely that they would *only* use specifications with covariates if covariates were only being used to improve precision.

Table 2: TSLS papers with covariates by journal and empirical specification

	(1)	(2)	(3)	(4)
		At least one specification		
	Papers using TSLS with covariates	Saturated in covariates	Not saturated in covariates	Saturated in instruments and covariates
American	100%	0%	100%	0%
Economic Review	36	0	36	0
Quarterly Journal	100%	4%	100%	0%
of Economics	24	1	24	0
Journal of Political	100%	16%	100%	0%
Economy	19	3	19	0
Econometrica	100%	9%	91%	9%
	11	1	10	1
Review of	100%	0%	100%	0%
Economic Studies	9	0	9	0
All	100 %	5%	99%	1%
	99	5	98	1

Notes: This table classifies the papers from column (3) of Table 1 by TSLS specification.

The saturate and weight (SW) specification (Angrist and Pischke, 2009) is saturated in covariates, and thus has rich covariates. However, it also uses a first stage that is fully saturated in both the covariates *and* the instruments, meaning that the regressors are indicators for all possible covariate-instrument combinations. This results in a large number of excluded variables and potential many instruments bias, which may explain why the SW specification was used by only a single paper in the survey. In fact, that one paper (Chamberlain and Imbens, 2004) is a methodological consideration of many instruments bias.

However, our results on monotonicity-correctness show that the interactions between covariates and instruments used in the SW specification may not be necessary for the LATE interpretation. Excluded interactions were not used in (1) and yet Propositions 1 and 2 show that if covariates are rich, then β_{iv} will be composed of only positively weighted complier effects. The reason is that the Angrist and Imbens (1995) monotonicity condition was *assumed* to work in the same direction for every covariate group. In contrast, the SW specification is premised on a version of the monotonicity assumption that allows the direction of monotonicity to vary with covariates. If one strengthens monotonicity—which we argue in Section 4.4 is often reasonable—then instrument-covariate interactions are not needed in the first stage to preserve the complier interpretation, at least if the instrument is scalar and ordered (see Proposition 10 ahead).

Our findings show that flexibly controlling for covariates is important for ensuring TSLS has a causal interpretation, but that these covariates can often enter separably from both the instruments and treatment. If a flexible covariate specification cannot be used, then another response to our findings is to test the null hypothesis that $\mathbb{L}[Z|X] = \mathbb{E}[Z|X]$. The most well-known test is Ramsey's (1969) RESET test, which is straightforward to implement (see, e.g. Wooldridge, 2010, pp. 137–138). If Z is binary, then it is also a good idea to check that the fitted values $\mathbb{L}[Z|X]$ lie between 0 and 1, which is necessary for $\mathbb{L}[Z|X] = \mathbb{E}[Z|X]$. Alternatively, researchers might consider using a method other than TSLS that is explicitly designed to estimate a LATE-type parameter, such as one of those cited in the introduction.

One such alternative method for the binary treatment, binary instrument case is the weighting approach developed by Abadie (2003). His approach requires estimating $\mathbb{E}[Z|X = x]$ in a first step, and thus makes explicit the implicit parametric assumption invoked when interpreting TSLS as LATE. Angrist and Pischke (2009, pp. 180–181) use Angrist's (2001) reanalysis of Angrist and Evans (1998) as an example to dismiss the relevance of Abadie's (2003) approach. Yet Angrist (2001, pg. 12) also reports that "the covariates are not highly correlated with the twins instruments. . . " Our findings show why it is misleading to extrapolate the Angrist and Pischke (2009) argument to other empirical settings: the case when Z is mean independent of X is one where *any* covariate specification is rich. If Z and X are dependent—as is usually the case when covariates are used in an IV analysis—then IV will not have a complier interpretation unless $\mathbb{E}[Z|X = x]$ is modeled correctly.

3 Definition and characterization of weakly causal estimands

In this section we define a weak property that an estimand should satisfy in order to "be causal." We interpret the estimands through a widely studied nonparametric IV model (e.g. Manski, 1989, 1994; Heckman, 1990; Imbens and Angrist, 1994; Balke and Pearl, 1994, 1997; Vytlacil, 2002, 2006). The model uses the potential outcomes notation with covariates used previously by Angrist and Imbens (1995), Angrist et al. (1996), Heckman and Vytlacil (1999), Angrist et al. (2000), Hirano et al. (2000), Abadie (2003), Tan (2006), Frölich (2007), Kolesár (2013), and Słoczyński (2022), among others. The model nests random assignment and selection on observables as special cases.

3.1 The nonparametric instrumental variables model

A discrete, ordered treatment variable T takes values in $\mathcal{T} \equiv \{t_0, t_1, \ldots, t_J\}$, listed in increasing order. We are interested in the causal effects that T has on an outcome variable, Y. We observe a scalar- or vector-valued instrumental variable (IV) Z that takes values in a set $\mathcal{Z}$. The case in Section 2 corresponds to $\mathcal{T} = \{0, 1\}$ and $\mathcal{Z} = \{0, 1\}$. There is a vector of covariates X with support $\mathcal{X}$.

Associated with each level of the IV is a potential treatment choice, $T(z)$. Associated with each level of the treatment is a potential outcome, $Y(t)$, which does not directly depend on the instrument due to the usual exclusion restriction. The potential and actual treatments and outcomes are related through

$$T = \sum_{z \in \mathcal{Z}} \mathbb{1}[Z = z]T(z) \quad \text{and} \quad Y = \sum_{t \in \mathcal{T}} \mathbb{1}[T = t]Y(t).$$

We maintain the following standard nonparametric exogeneity condition throughout our analysis.

Assumption EX. (Exogeneity) $(\{T(z)\}_{z \in \mathcal{Z}}, \{Y(t)\}_{t \in \mathcal{T}}) \perp\!\!\!\perp Z | X$.

We assume that each of T, Z, and X are discretely distributed with finite support. This is just for mathematical simplicity. Our theoretical results can be extended to allow for T to be a continuous scalar, and both X and Z to be vectors with continuous components. The changes required essentially involve replacing sums with integrals and finite indices with function arguments. We also assume throughout that the expectation of Y exists.

Our analysis is based on partitioning individuals into mutually exclusive and exhaustive groups based on their potential treatment choices. Order $\mathcal{Z}$ arbitrarily as $\mathcal{Z} \equiv \{z_0, z_1, \ldots, z_K\}$. Let $G \equiv (T(z_0), T(z_1), \ldots, T(z_K))$ denote an individual's choice group, that is, their configuration of potential treatment choices under each of the instrument values. Let $\mathcal{G}$ denote the values that G can take. In the binary treatment ($\mathcal{T} = \{0, 1\}$), binary instrument ($\mathcal{Z} = \{0, 1\}$) case, G takes values in $\mathcal{G} = \{(0, 0), (1, 1), (0, 1), (1, 0)\}$, corresponding to the groups Angrist et al. (1996, Table 1) called the never-takers, always-takers, compliers, and defiers, respectively. Using the group notation, Assumption EX can be equivalently written as follows.

Assumption EX. (Exogeneity, group form) $(G, \{Y(t)\}_{t \in \mathcal{T}}) \perp\!\!\!\perp Z | X$.

3.2 Definition of a weakly causal estimand

Consider the *group treatment responses* (GTRs)

$$\mu_j(g,x) \equiv \mathbb{E}[Y(t_j)|G=g, X=x],$$

which are the expected potential outcomes across choice and covariate groups.[8] We collect the GTRs as $\mu \equiv \{\mu_j(g,x) : j = 0, 1, \ldots, J, g \in \mathcal{G}, x \in \mathcal{X}\}$, which takes values in some set $\mathcal{M} \subseteq \mathbb{R}^{d_\mu}$ that reflects any additional maintained assumptions on the GTRs, such as Assumption CLE ahead. We use the following definition as a minimal requirement for an estimand to be interpreted as "causal."

Definition WC. β is *weakly causal* if both of the following statements are true for all $\mu \in \mathcal{M}$:

$$\begin{aligned}
&\text{If } \mu_j(g,x) - \mu_{j-1}(g,x) \geq 0 \text{ for all } j \geq 1, \text{ all } g \in \mathcal{G}, \text{ and every } x \in X, \text{ then } \beta \geq 0.\\
&\text{If } \mu_j(g,x) - \mu_{j-1}(g,x) \leq 0 \text{ for all } j \geq 1, \text{ all } g \in \mathcal{G}, \text{ and every } x \in X, \text{ then } \beta \leq 0. \quad (4)
\end{aligned}$$

Definition WC is a natural requirement for an estimand β to reflect the causal effect of T on Y. The requirement is merely that *if* the causal effect of the treatment has the same sign for every treatment contrast, and every choice and covariate subgroup, then the summary estimand β also has that sign. That is, β is weakly causal if it is not a *systematically* misleading measure of the sign of the underlying group- and covariate-specific treatment effects.

Definition WC is intended to be an extremely weak criterion. An estimand can be weakly causal and still be completely uninteresting. For example, the trivial estimand $\beta = 0$ is weakly causal. However, it seems unlikely that an estimand that fails to be weakly causal could still reasonably be described as reflecting the causal effect of T on Y, since it may not even have the right sign. As minimal as Definition WC is, we have already seen in Figure 1 that a linear IV estimand can fail to satisfy it, even if the instrument satisfies exclusion and exogeneity (Assumption EX).

We provide sufficient and necessary conditions to be weakly causal for any estimand that can be written as

$$\beta = \mathbb{E}[b(T,X,Z)Y] \tag{5}$$

for some function b. For example, β_{iv} in Section 2 satisfies (5) with $b(T,X,Z) =$

[8] As a minor abuse of notation, we assume that $\mu_j(g,x)$ is well-defined for all (g,x), even if g is not in the support of G given X, so that $\mathbb{P}[G=g, X=x] = 0$. This convention has no impact on our results.

$\tilde{Z}/\mathbb{E}[T\tilde{Z}] = (Z - \mathbb{L}[Z|X])/\mathbb{E}[T(Z - \mathbb{L}[Z|X])]$. The following result decomposes these estimands into GTRs.

Proposition 3. Suppose that β has form (5), and that Assumption EX holds. Then

$$\beta = \sum_{g,x} \omega_0(g,x)\mu_0(g,x) + \sum_{g,x}\sum_{j=1}^{J} \omega_j(g,x)\left(\mu_j(g,x) - \mu_{j-1}(g,x)\right), \tag{6}$$

where $\omega_j(g,x) \equiv \mathbb{E}\left[\mathbb{1}[T \geq t_j]b(t_j,x,Z)|G=g,X=x\right]\mathbb{P}[G=g,X=x]$ for all $j \geq 0$.

3.3 Weak causality and positively weighted averages

Suppose that μ is only restricted to lie in some hypercube $\mathcal{M}_\Box \equiv [\underline{y},\overline{y}]^{d_\mu}$, where $\underline{y} < \overline{y}$ and either or both of $\underline{y}$ and $\overline{y}$ could be infinite. Then Definition WC is equivalent to the widely-used criterion of positively weighted subgroup-specific treatment effects (e.g. Angrist, 1998; Lee, 2008a; Angrist and Pischke, 2009; Card et al., 2015; Goodman-Bacon, 2021; Sun and Abraham, 2021; Goldsmith-Pinkham et al., 2021).

Proposition 4. Suppose that β has the form (5), that Assumption EX holds, and that $\mathcal{M} = \mathcal{M}_\Box$. Then β is weakly causal if and only if:

- **(Non-negative weights)** $\omega_j(g,x) \geq 0$ for all $j \geq 1$, and all g and x.
- **(Level irrelevance)** $\omega_0(g,x) = 0$ for all g and x.

When these conditions are satisfied,

$$\beta = \sum_{g,x}\sum_{j=1}^{J} \omega_j(g,x)\left(\mu_j(g,x) - \mu_{j-1}(g,x)\right) \tag{7}$$

for non-negative weights $\omega_j(g,x) \geq 0$.

Proposition 3 shows that β can always be written as (6). Proposition 4 uses that representation to show that if β cannot also be written like (7) with weights that are non-negative, then one of two things must be true: either β only reflects treatment effects, but some of these effects are negatively weighted, or else β reflects not just treatment *effects* but also the *levels* of potential outcomes. The first situation violates the non-negative weights requirement, which is naturally necessary for β to be weakly causal (recall Figure 1). The second situation violates the level irrelevance requirement. Level irrelevance is necessary for β to be weakly causal because it prevents the possibility that all treatment *effects* are positive, even while the *levels* of the GTRs are such that $\beta < 0$.

The necessary direction of Proposition 4 can be altered by further restricting $\mathcal{M}$ so that the GTRs must satisfy additional properties. The most salient assumption is that treatment effects are constant and linear.

Assumption CLE. (Constant, linear effects) There exists a constant Δ such that $\mu_j(g, x) - \mu_{j-1}(g, x) = \Delta(t_j - t_{j-1})$ for every $j \geq 1$, $g \in \mathcal{G}$ and $x \in \mathcal{X}$.

We let $\mathcal{M}_{\text{CLE}}$ denote the subset of $\mathcal{M}_{\square}$ that also satisfies the condition in Assumption CLE. The next proposition shows that with constant, linear treatment effects, the requirement of non-negative weights can be relaxed so that the non-negativity happens after aggregating across groups, covariates, and treatment contrasts.

Proposition 5. Suppose that β has the form (5) and that Assumptions EX and CLE hold, so that $\mathcal{M} = \mathcal{M}_{\text{CLE}}$. Then β is weakly causal if and only if:

- **(Non-negative weights, aggregated)** $\sum_{g,x} \sum_{j=1}^{J} (t_j - t_{j-1}) \omega_j(g, x) \geq 0$.
- **(Level irrelevance)** $\omega_0(g, x) = 0$ for all g and x.

4 When is TSLS weakly causal?

In this section we specialize the general results of the previous section to a large class of TSLS estimands.

4.1 TSLS specifications and estimands

A TSLS specification is characterized by four components: (i) the outcome variable; (ii) the variables included in the second stage, but not the first; (iii) the variables included in the first stage but excluded from the second; and (iv) the variables included in both stages. The nonparametric IV model specifies the outcome variable, Y, but not which combinations of T, Z, and X go in the first and second stages. We consider TSLS specifications where (ii) is the treatment, T, (iii) is a vector of instruments $I \equiv i(Z, X)$, where i is a known, vector-valued function, and (iv) is a vector of covariates, $C \equiv c(X)$, where c is also a known, vector-valued function. Together, $F \equiv [I', C']'$ are the first stage variables, while $S \equiv [T, C']'$ are the second stage variables.

One way to interpret the first stage of TSLS is as a procedure for reducing F down to the same dimension as S by transforming I into a scalar. That is, the first stage of TSLS replaces the vector of instruments I by the scalar *effective instrument*

$$\dot{Z} \equiv \gamma' I,$$

where γ is the vector of population coefficients on I in the first stage regression of T on I and C. The TSLS estimand can then be written as the standard IV estimand that uses $\dot{F} \equiv [\dot{Z}, C']'$ as instruments for $S \equiv [T, C']'$, that is

$$\alpha_{\text{tsls}} = \mathbb{E}[\dot{F}S']^{-1}\,\mathbb{E}[\dot{F}Y]. \tag{8}$$

Alternatively, the first stage of TSLS can be viewed as constructing fitted values for the treatment,

$$\dot{T} \equiv \dot{Z} + \lambda' C,$$

where λ is the vector of population coefficients on C in the first stage regression. The TSLS estimand is then the OLS estimand from a regression of Y onto $\dot{T}$ and C.[9]

We assume throughout that the standard rank condition holds, so that α_{tsls} exists. Our interest is in the component of α_{tsls} that corresponds to the coefficient on T, since it is this coefficient that could potentially be used to measure the causal effect of T on Y. We call this component β_{tsls}. An expression for β_{tsls} like (5) can be found by applying the Frisch-Waugh-Lovell Theorem.

Proposition 6. Let β_{tsls} denote the component of α_{tsls} that corresponds to the coefficient on T. Then

$$\beta_{\text{tsls}} = \frac{\mathbb{E}[\tilde{Z}Y]}{\mathbb{E}[\tilde{Z}^2]} = \frac{\mathbb{E}[\tilde{Z}Y]}{\mathbb{E}[\tilde{Z}T]} = \mathbb{E}\left[\left(\frac{\tilde{Z}}{\mathbb{E}[\tilde{Z}T]}\right)Y\right],$$

where $\tilde{Z} \equiv \dot{Z} - \mathbb{L}[\dot{Z}|C]$ are the population residuals from regressing $\dot{Z}$ onto C and $\mathbb{L}[\dot{Z}|C] \equiv \mathbb{E}[\dot{Z}C']\,\mathbb{E}[CC']^{-1}C$ are the population fitted values.

Propositions 3 and 6 imply that β_{tsls} can be written as (6) with

$$\omega_j(g,x) = \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}\left[\mathbb{1}[T \geq t_j]\tilde{Z}|G = g, X = x\right]\mathbb{P}[G = g, X = x]. \tag{9}$$

As shown in Propositions 4 and 5, whether β_{tsls} is weakly causal is determined by $\omega_j(g,x)$. The properties of $\omega_j(g,x)$ are in turn determined by the TSLS specification. Next, we study how aspects of the TSLS specification affect whether $\omega_j(g,x)$ satisfies the conditions in Propositions 4 and 5, starting with the simpler case of constant, linear treatment effects considered in Proposition 5.

[9] Our definition of the TSLS estimand presumes the standard asymptotic framework where the number of observations is growing and the dimensions of I and C are fixed. Kolesár (2013) and Evdokimov and Kolesár (2019) consider alternative frameworks that allow for the dimensions of either or both of these vectors to also be growing.

4.2 Constant, linear treatment effects

In this section we assume constant, linear treatment effects (Assumption CLE).

4.2.1 Sufficient and necessary condition for weak causality

By itself, Assumption CLE does not guarantee that β_{tsls} is weakly causal. The sufficient and necessary condition is that the covariate specification is rich enough to reproduce the conditional mean of the effective instruments.

Definition RC. Let $\mathbb{L}[\dot{Z}|C = c(x)] \equiv \mathbb{E}[\dot{Z}C'] \, \mathbb{E}[CC']^{-1}c(x)$ be the population fitted value at $C = c(x)$ from regressing $\dot{Z}$ onto C. A TSLS specification has *rich covariates* if $\mathbb{E}[\dot{Z}|X = x] = \mathbb{L}[\dot{Z}|C = c(x)]$ for every $x \in \mathcal{X}$.

Proposition 7. Suppose that Assumptions EX and CLE are satisfied and that C contains a constant regressor. Then β_{tsls} is weakly causal if and only if the TSLS specification has rich covariates.

The intuition behind Proposition 7 is that a TSLS specification that does not have rich covariates reflects not just treatment *effects*, but also the *levels* of potential outcomes. For example, when $\mathcal{T} = \{0, 1\}$ with $Y = Y(0) + \Delta T$, Proposition 6 implies that

$$\beta_{\text{tsls}} = \mathbb{E}[\tilde{Z}T]^{-1} \, \mathbb{E}[\tilde{Z}(Y(0) + \Delta T)] = \Delta + \overbrace{\mathbb{E}[\tilde{Z}T]^{-1} \, \mathbb{E}[\tilde{Z}Y(0)]}^{\text{depends on } Y(0)}. \tag{10}$$

Using Assumption EX, the potentially level-dependent term can be written as

$$\mathbb{E}[\tilde{Z}Y(0)] = \mathbb{E}\left[\mathbb{E}[\tilde{Z}|X] \, \mathbb{E}[Y(0)|X]\right]. \tag{11}$$

The nonparametric IV model does not restrict $\mathbb{E}[Y(0)|X]$ at all. Level-dependence will thus happen whenever $\mathbb{E}[\tilde{Z}|X] \neq 0$ with positive probability, which in turn happens whenever the TSLS specification does not have rich covariates, because

$$\mathbb{E}[\tilde{Z}|X] \equiv \mathbb{E}[\dot{Z}|X] - \mathbb{L}[\dot{Z}|C].$$

An alternative way to interpret this level-dependence is as an asymmetric weighting of $Y(0)$ and $Y(1)$. Using the expression in Proposition 6,

$$\beta_{\text{tsls}} = \mathbb{E}\left[\mathbb{E}[\tilde{Z}T]^{-1}\tilde{Z}TY(1)\right] - \mathbb{E}\left[\mathbb{E}[\tilde{Z}T]^{-1}\left(\tilde{Z}T - \tilde{Z}\right)Y(0)\right], \tag{12}$$

which shows that β_{tsls} is the difference between weighted averages of $Y(1)$ and $Y(0)$. The weights in these averages are not in general the same—the weights are asymmetric—because the weight on $Y(0)$ also includes the term $-\mathbb{E}[\tilde{Z}T]^{-1}\tilde{Z}$. If the covariates are rich, then this additional term will always be uncorrelated with $Y(0)$, and the weights in (12) return to being symmetric.

4.2.2 Discussion

The assumption that a TSLS specification has rich covariates has been used previously in the literature. Abadie (2003, Proposition 5.2) proved the sufficient direction of Proposition 7 in the special case that $\mathcal{T} = \{0, 1\}$, $I = Z$ and $\mathcal{Z} = \{0, 1\}$. Kolesár (2013, pp. 8–10) proves the sufficient direction for the more general case considered here.

Proposition 7 shows that assuming rich covariates is not only sufficient for β_{tsls} to have a causal interpretation, it is also *necessary.* Whenever a TSLS specification does not have rich covariates, β_{tsls} will not be weakly causal. Moreover, the necessity has nothing to do with heterogeneous or nonlinear treatment effects. Rather, it is a fundamental consequence of the exercise started by Imbens and Angrist (1994) of interpreting a linear IV estimand through a nonparametric IV model. The linear IV estimator was designed for the linear IV model; giving it a causal interpretation within a nonparametric IV model requires additional assumptions.

As Kolesár (2013, pp. 10–11) notes, there are two important special cases in which a TSLS specification will have rich covariates. One is when X is discrete and C contains an indicator for each possible realization of X, so that the specification is saturated in X. The other is when Z is unconditionally randomly assigned, and I contains only functions of Z, so that $\mathbb{E}[\dot{Z}|X = x] = \gamma' \mathbb{E}[I|X = x] = \gamma' \mathbb{E}[i(Z)]$ is constant in x. Outside of these two special cases, the claim that a TSLS specification has rich covariates is one that must be defended. Using domain knowledge to argue that $\mathbb{E}[\dot{Z}|X = x]$ is indeed linear in $c(x)$ seems difficult, and indeed we found no papers in our survey that tried to do so.

Proposition 7 also applies to selection on observables by taking $Z = T = I$. Angrist (1998) proposed implementing a selection on observables strategy using a linear regression with a single binary treatment indicator and saturated covariates, describing the difference between this regression and nonparametric matching as "partly cosmetic" (Angrist, 1998, pg. 255). Proposition 7 shows that Angrist's (1998) argument cannot be extrapolated beyond the saturated case: any deviation from full saturation will lead to negative weights, at least without an auxiliary functional form assumption.[10] Moreover, whenever Angrist's (1998) saturated specification can actually be implemented, the overlap condition $\mathbb{P}[T = 1|X = x] \in (0, 1)$ must hold for every x, or else there would be perfect multicollinearity. Thus, a linear regression implementation of selection on observables is weakly causal only when it is also possible to estimate a traditional parameter (such as the average treatment on the treated) using a nonparametric matching estimator.[11]

4.2.3 Imposing an additional linearity assumption

One way to overturn the necessary direction of Proposition 7 is to add the assumption that $\mathbb{E}[Y(t_j)|X = x]$ is a linear function of $c(x)$.

Assumption LIN. (Linear potential outcome mean) $\mathbb{E}[Y(t_j)|X = x] = \eta' c(x)$ for some η and some j.

Proposition 8. Suppose that Assumptions EX, CLE, and LIN are satisfied, and that C contains a constant regressor.[12] Then $\beta_{\text{tsls}} = \Delta$, so β_{tsls} is weakly causal.

Assumption LIN—or something similar—is explicitly stated in classical and textbook treatments of IV models, e.g. Heckman and Robb (1985, pp. 184–186) or Wooldridge (2010, pg. 939). But it is not part of the nonparametric IV model on which the widely-invoked "LATE interpretation" of TSLS rests (Angrist and Imbens, 1995). In our survey, we found no researchers who attempted to justify an assumption like Assumption LIN. As Abadie (2003, pg. 247) points out, an undesirable implication of Assumption LIN is that one can have $\beta_{\text{tsls}} = \Delta$ even if the excluded "instruments" I depend only on X, and not on Z, an example of what Angrist and Pischke (2009, pg. 191) describe as "back-door identification."

A higher-level alternative to having rich covariates or imposing Assumption LIN is to directly assume that the left-hand side of (11) is zero. This assumption appears in Wooldridge's (2010, pg. 937) discussion of the binary treatment case as the assumption that $\mathbb{L}[Y(0)|C, I]$ does not depend on I. If we put aside knife-edge balancing cases, (11) shows that this assumption either requires rich covariates or Assumption LIN. However,

[10] Assumption LIN in the next section is one such assumption. Under this assumption, selection on observables with a binary treatment is a special case of the analysis in Section 2 in which every unit is a complier. Assumption LIN is, however, crucial for this conclusion.

[11] There may however be statistical differences between such estimators. Goldsmith-Pinkham et al. (2021) argue that the linear regression implementation will necessarily have a smaller asymptotic variance.

[12] In terms of the $\mathcal{M}$ set in Section 3, $\mathcal{M} = \mathcal{M}_{\text{CLE}} \cap \mathcal{M}_{\text{LIN}}$, where $\mathcal{M}_{\text{LIN}} \equiv \{\mu : \sum_g \mu_j(g, x)\,\mathbb{P}[G = g|X = x] = \eta' c(x)$ for some η and some $j\}$.

considering the high-level assumption usefully exposes the fundamental problem with using the nonparametric IV model to justify TSLS: Assumption EX by itself *does not* imply that regressing $Y(0)$ onto C and I would yield a zero coefficient on I, even though this condition is essential for giving TSLS a causal interpretation.

Assumption LIN was also maintained in Propositions 1 and 2, which showed that β_{tsls} is not weakly causal without rich covariates. This does not contradict Proposition 8 because of the addition of constant, linear treatment effects (Assumption CLE). When Assumption CLE is removed to allow for heterogeneous treatment effects, Assumption LIN no longer suffices as a substitute for rich covariates.

4.3 Heterogeneous treatment effects

In this section we drop Assumption CLE and allow for treatment effects that are both nonlinear and heterogeneous across both covariates and groups, as in Angrist and Imbens (1995).

4.3.1 Monotonicity

We add a conditional-on-covariates form of the Imbens and Angrist (1994) monotonicity condition. We follow Słoczyński (2022) in calling this "weak" monotonicity, in contrast to "strong" monotonicity, introduced ahead.

Assumption WM. (Weak monotonicity) For all $x \in \mathcal{X}$, and all $z, \bar{z} \in \mathcal{Z}$, either

$$\begin{aligned} & \mathbb{P}[T(\bar{z}) \geq T(z) | X = x] = 1 \\ \text{or} \qquad & \mathbb{P}[T(z) \geq T(\bar{z}) | X = x] = 1. \end{aligned}$$

We describe Assumption WM as *weak* monotonicity because it allows the direction of monotonicity to depend on x. For example, if $\mathcal{Z} = \{0, 1\}$ and $\mathcal{X} = \{0, 1\}$, then Assumption WM allows for

$$\begin{aligned} & \mathbb{P}[T(1) \geq T(0) | X = 0] = 1 \\ \text{and} \qquad & \mathbb{P}[T(0) \geq T(1) | X = 1] = 1. \end{aligned} \tag{13}$$

If, for example, $\mathcal{T} = \{0, 1\}$ is also binary, then group $G = (0, 1)$ would be compliers conditional on $X = 0$, but they would be defiers conditional on $X = 1$, and conversely for $G = (1, 0)$.

4.3.2 Sufficient and necessary conditions for weak causality

For any x, the order in which Assumption WM holds between two instrument values can be determined by the conditional mean of T,

$$p(z, x) \equiv \mathbb{E}[T|Z = z, X = x].$$

If $p(\bar{z}, x) \geq p(z, x)$ then $T(\bar{z}) \geq T(z)$ conditional on $X = x$, and conversely (Imbens and Angrist, 1994; Vytlacil, 2002). We say that the first stage of the TSLS specification is monotonicity-correct if the first stage fitted values reproduce this ordering, in the sense of predicting higher values of treatment when the instrument is such that individuals choose higher values of treatment.

Definition MC. Let $\dot{t}(z, x) \equiv \gamma' i(z, x) + \lambda' c(x)$ denote the population fitted values in the first stage regression for a realization with $Z = z$ and $X = x$. Suppose that $(z, \bar{z})$ are both in the support of Z, conditional on $X = x$. Then a TSLS first stage is *monotonicity-correct* for $(z, \bar{z})$ conditional on $X = x$, if

$$(p(\bar{z}, x) - p(z, x)) \times \left(\dot{t}(\bar{z}, x) - \dot{t}(z, x)\right) \geq 0.$$

In the previous section, we established that the requirement of rich covariates is necessary for β_{tsls} to be weakly causal. Since this was true under Assumption CLE, it also remains true without that assumption. (And, as already noted, it also remains true even if one imposes Assumption LIN, as shown in Proposition 2.) With heterogeneous treatment effects, and given rich covariates, whether β_{tsls} is weakly causal depends on the monotonicity-correctness of its first stage.

Proposition 9. Suppose that Assumptions EX and WM are satisfied. Suppose that the TSLS specification for β_{tsls} has rich covariates and that C contains a constant regressor. If the TSLS specification is monotonicity-correct for every $(z, \bar{z})$, conditional on every x, then β_{tsls} is weakly causal. Conversely, if β_{tsls} is weakly causal, then for every $x \in \mathcal{X}$ the TSLS first stage must be monotonicity-correct for at least one pair $(z, \bar{z})$.

4.3.3 Interpreting monotonicity-correctness

Definition MC is easiest to appreciate in the case with $\mathcal{T} = \{0, 1\}$, $I = Z$, and $Z \in \{0, 1\}$, so that $\dot{t}(1, x) - \dot{t}(0, x) = \gamma$ is the scalar coefficient on Z in the first stage regression. If Assumption WM holds with $T(1) \geq T(0)$ conditional on $X = x$, then $p(1, x) - p(0, x) \geq 0$. The TSLS first stage is monotonicity-correct conditional on $X = x$

if and only if $\gamma > 0$, so that the linear projection in the first stage reproduces the same sign as the (nonparametric) propensity score.

Proposition 9 shows that a *necessary* condition for β_{tsls} to be weakly causal in this case is that the TSLS first stage is monotonicity-correct *for all x*. Since there is only a single coefficient γ, the implication is that β_{tsls} for this specification will not be weakly causal if the direction of Assumption WM changes with x, as in (13). Słoczyński (2022, Theorem 3.3) previously made this point in the binary Z, binary T case with C saturated in X.

Including interactions between covariates and instruments in the first stage can help ensure monotonicity-correctness. For example, suppose that X contains a binary component, $X_1 \in \{0, 1\}$, and that $I = [Z, ZX_1]'$ now has two components with first stage coefficient vector $[\gamma_1, \gamma_2]'$, so that for any realization of the other components x_{-1} of X,

$$\begin{aligned} &\dot{t}(1, x_1 = 0, x_{-1}) - \dot{t}(0, x_1 = 0, x_{-1}) = \gamma_1 \\ \text{and} \quad &\dot{t}(1, x_1 = 1, x_{-1}) - \dot{t}(0, x_1 = 1, x_{-1}) = \gamma_1 + \gamma_2. \end{aligned}$$

This first stage can still be monotonicity-correct conditional on all values of $X = (x_1, x_{-1})$, even if the direction of Assumption WM is positive when $x_1 = 0$ and negative when $x_1 = 1$, as in (13). The requirement is that $\gamma_1 \geq 0$ and $\gamma_1 + \gamma_2 \leq 0$. Whether this requirement holds depends on how the covariates C are specified, and on the stochastic relationship between Z and the other components, X_{-1} (see Section 4.4).

As this example suggests, monotonicity-correctness is about how the relationship between Z and T varies conditional on X. This is different from rich covariates, which is about the joint distribution of Z and X. It is possible for a TSLS specification to be monotonicity-incorrect even if Z and X are independent, as in a completely randomized experiment.

With a binary instrument $\mathcal{Z} = \{0, 1\}$, the sufficient and necessary conditions in Proposition 9 are the same. With a multivalued instrument, a small gap opens up between the two conditions. The gap occurs because it is possible—at least in principle—for the first stage to be monotonicity-incorrect for some instrument contrasts, as long as it is monotonicity-correct "on average" across all instrument contrasts. This type of fortuitous averaging seems difficult to defend, so for practical purposes we view the gap between sufficient and necessary in Proposition 9 as empirically irrelevant.

4.3.4 Relationship to the literature

Special cases of the sufficient conditions in Proposition 9 appear in Angrist and Imbens (1995), Angrist and Pischke (2009), Kolesár (2013), and Słoczyński (2022). Angrist and Imbens (1995) and Angrist and Pischke (2009) assume first stage specifications that are saturated in both the instruments and covariates, which is automatically monotonicity-correct for any instrument pair, conditional on any covariate value. Kolesár (2013) relaxes this to Definition MC, although stated somewhat differently; see also Heckman and Vytlacil (2005, Section 4.3) and Heckman (2010, Section 3.4).

These authors primarily consider the binary treatment case. For multivalued treatments, Angrist and Imbens (1995), Angrist and Pischke (2009), and Kolesár (2013) interpret β_{tsls} as a positively weighted average of "average causal response" functions, rather than of underlying subgroup-specific treatment effects. Heckman et al. (2006, pp. 414–415) provide an interpretation in terms of subgroup-specific treatment effects when the first stage is fully saturated, but under a condition different from Assumption WM.

In contrast, Proposition 9 shows that—given rich covariates—having a monotonicity-correct first stage is not only sufficient for a causal interpretation, it is also necessary. It is not just necessary for interpreting β_{tsls} as a positively weighted average of LATEs, but even for interpreting β_{tsls} as weakly causal. Proposition 9 thus exactly characterizes the set of TSLS specifications that can be said to have a causal interpretation under Assumptions EX and WM. This set turns out to be essentially the same as the restrictive fully saturated one originally used in Angrist and Imbens (1995, Theorem 3), the one which was used in only a single one of the 99 papers in our survey that used TSLS with covariates (Section 2.4).

4.4 Which TSLS estimands are weakly causal?

Propositions 8 and 9 show that for a TSLS specification to produce a weakly causal estimand it needs to be both rich and monotonicity-correct.

Nonparametric TSLS specifications that restrict attention to the population with $X = x$ and use a first stage that is saturated in Z will be both rich (trivially) and monotonicity-correct. Each value of x produces a different estimand $\beta_{\text{tsls}}(x)$, each of which is weakly causal. Any positively weighted sum of $\beta_{\text{tsls}}(x)$ across $x \in \mathcal{X}$ will be weakly causal. Frölich (2007) discusses several different weighting schemes.

Nonparametric conditioning can be viewed as arising from a TSLS specification that fully interacts both the treatment and instruments with indicators for each x bin. The "saturate and weight" (SW) specification (Angrist and Pischke, 2009; Angrist and Imbens, 1995) is like nonparametric conditioning but uses only a single treatment variable. Letting $\mathcal{X} \equiv \{x_1, \dots, x_L\}$, the SW specification takes

$$\begin{aligned} & C \equiv [1, \mathbb{1}[X = x_\ell] : \ell = 2, \dots, L]' \qquad \text{(SW)} \\ \text{and} \quad & I \equiv [\mathbb{1}[Z = z_k], \mathbb{1}[X = x_\ell]\mathbb{1}[Z = z_k] : \ell = 2, \dots, L, \text{ and } k = 2, \dots, K]'. \end{aligned}$$

Specification SW is both rich and monotonicity-correct: it is rich because regressing Z onto C yields fitted values $\mathbb{E}[Z|X]$, and it is monotonicity-correct because the first stage fitted values are equal to the propensity score.

However, because specification SW has L excluded variables—the same number of values that X takes—it is quite vulnerable to many instruments bias. Jackknife estimators have been proposed for reducing many instruments bias (Angrist et al., 1999; Ackerberg and Devereux, 2009; Kolesár, 2013). In Appendix B, we report Monte Carlo evidence on the performance of the JIVE, IJIVE, and UJIVE estimators for specification SW. Our results confirm that many instruments bias can be a serious problem for estimating specification SW with TSLS, and often remains a substantial problem even when using jackknife estimators.

The large number of excluded variables in specification SW are created by interacting X and Z. Removing these interactions produces the saturated separable (SS) specification

$$C \equiv [1, \mathbb{1}[X = x_\ell] : \ell = 2, \dots, L]' \quad \text{and} \quad I \equiv [\mathbb{1}[Z = z_k] : k = 2, \dots, K]'. \qquad \text{(SS)}$$

Specification SS has the same covariates C as specification SW, so it still has rich covariates. However, removing the interactions means that specification SS will not be monotonicity-correct if the direction of monotonicity changes with X, as in (13).

A natural response is to strengthen Assumption WM to require the direction of monotonicity to be invariant to X. Słoczyński (2022) calls this strong monotonicity.

Assumption SM. (Strong monotonicity) For all $z, \bar{z} \in \mathcal{Z}$, either

$$\begin{aligned} & \mathbb{P}[T(\bar{z}) \geq T(z)|X = x] = 1 \\ \text{or} \qquad & \mathbb{P}[T(z) \geq T(\bar{z})|X = x] = 1 \quad \text{for } \textit{all } x. \end{aligned}$$

Assumption SM implies that $p(\bar{z}, x) - p(z, x)$ has the same sign for all x. Perhaps surprisingly, however, specification SS is still not necessarily monotonicity-correct even under Assumption SM. The reason is that omitted interaction terms can bias the coefficients on the instrument indicators in a way that contradicts the sign of the propensity score (see Appendix C for an example).

Suppose, however, that Z is scalar and ordered, and that the comparisons in Assumption SM follow this ordering.

Assumption OSM. (Ordered strong monotonicity) Z is scalar with $z_0 \leq z_1 \leq \cdots \leq z_K$ and

$$\mathbb{P}[T(z_0) \leq T(z_1) \leq \cdots \leq T(z_K)|X = x] = 1 \quad \text{for all } x.$$

Assumption OSM also does not ensure that specification SS is monotonicity-correct. However, it does ensure that the more parsimonious saturated linear (SL) specification

$$C \equiv [1, \mathbb{1}[X = x_\ell] : \ell = 2, \ldots, L]' \quad \text{and} \quad I = Z. \tag{SL}$$

will be monotonicity-correct. More generally, any specification with rich covariates that has a first stage that is separable and linear in Z will produce a β_{tsls} that is weakly causal.

Proposition 10. Suppose that Assumptions EX and OSM are satisfied. If the TSLS specification has rich covariates with C containing a constant regressor, and if $I = Z$, then β_{tsls} is weakly causal.

These findings suggests that the original Angrist and Imbens (1995) monotonicity condition, Assumption WM, is often too weak to ensure reasonable TSLS estimands are weakly causal. On the other hand, it is not clear that strengthening Assumption WM to Assumption SM or OSM appreciably changes its economic content.[13]

For example, Angrist and Evans (1998) use parental preferences for mixed sibling sex composition as a binary instrument for fertility, with the argument that having two children of different sexes ($Z = 0$) makes a family less likely to have a third child ($T = 1$). Assumptions WM, SM, and OSM are all potentially suspicious in this case because they require all families to have the same sex-mix preferences, ruling out the existence of *any* family that prefers two boys, and thus eliminating meaningful unobserved heterogeneity in preferences.[14] The difference between Assumptions WM and SM (or OSM) is whether sex-mix preferences can be modulated by observables, e.g. more educated families *all* prefer mixed-sex, while less educated families *all* prefer two boys. While mathematically weaker, such an assumption does not address the critique

[13] Słoczyński (2022) provides a more positive view of Assumption WM relative to Assumption SM.

[14] Rosenzweig and Wolpin (2000) and Lee (2008b) show that such preferences are indeed relevant to the fertility behavior of parents in India and South Korea, respectively.

that there is unobserved heterogeneity in parental preferences for mixed sibling sex composition.

5 Decomposing TSLS estimands

A TSLS estimand that is not weakly causal could still be "close" to weakly causal if the covariates are "almost" rich and the first stage is "almost" monotonicity-correct. To make this concept precise, we develop a decomposition that breaks a TSLS estimand into a weakly causal term and two terms that measure the degree to which covariates are not rich and the first stage is not monotonicity-correct. All three of these terms are point identified, so we can use the decomposition to estimate the extent to which a TSLS estimand fails to be weakly causal.

For any x, let $\xi_0(x)$ denote the value of z for which $p(z, x)$ is smallest, then let $\xi_1(x)$ denote the second smallest value, $\xi_2(x)$ the third smallest, and so on. Then $p(\xi_0(x), x) \leq p(\xi_1(x), x) \leq \cdots \leq p(\xi_K(x), x)$, where K is the number of elements of $\mathcal{Z}$. For $k \geq 1$, let

$$\Upsilon_k(x) \equiv \mathbb{E}[Y|X = x, Z = \xi_k(x)] - \mathbb{E}[Y|X = x, Z = \xi_{k-1}(x)].$$

Also, let $\Xi_k(x) \equiv \{\xi_\ell(x)\}_{\ell \geq k}$ denote the set of instrument values with propensity score at least as large as $p(\xi_k(x), x)$, conditional on $X = x$. We use these definitions in the following proposition.

Proposition 11. Suppose that Assumption EX is satisfied. If a TSLS specification has rich covariates, then

$$\beta_{\text{tsls}} = \mathbb{E}\left[\sum_{k=1}^{K} \Upsilon_k(X)\tilde{t}_k(X)\phi_k(X)\right] \equiv \beta_{\text{rich}},$$

where

$$\begin{aligned}
\phi_k(x) &\equiv \mathbb{P}[Z \in \Xi_k(x)|X = x]\,\mathbb{P}[Z \notin \Xi_k(x)|X = x]\,\mathbb{E}[\tilde{Z}^2]^{-1} \\
\text{and} \quad \tilde{t}_k(x) &\equiv \mathbb{E}\left[\dot{t}(Z, x)|Z \in \Xi_k(x), X = x\right] - \mathbb{E}\left[\dot{t}(Z, x)|Z \notin \Xi_k(x), X = x\right].
\end{aligned}$$

Proposition 11 provides the following decomposition:

$$\beta_{\text{tsls}} = \beta_{\text{rich}} + (\beta_{\text{tsls}} - \beta_{\text{rich}}). \tag{14}$$

The first term, β_{rich}, is the estimand that would have been obtained by a given TSLS specification had the distribution of observables been such that the covariates were rich, i.e. such that $\mathbb{E}[\dot{Z}|X = x] = \mathbb{L}[\dot{Z}|C = c(x)]$ for all x. The second term is the deviation between the estimand produced under the actual distribution of the observables, β_{tsls}, and the estimand β_{rich} produced under this idealized scenario. Since each of $\Upsilon_k(x), \tilde{t}_k(x)$, and $\phi_k(x)$ are features of the distribution of observables, we can directly estimate both terms in (14).

With heterogeneous treatment effects, β_{rich} could still reflect both positively and negatively weighted treatment effects depending on the sign of $\tilde{t}_k(x)$. To see this, notice that each outcome contrast, $\Upsilon_k(x)$, is a numerator for a conditional Wald estimand. That is,

$$\text{WALD}_k(x) \equiv \frac{\mathbb{E}[Y|X = x, Z = \xi_k(x)] - \mathbb{E}[Y|X = x, Z = \xi_{k-1}(x)]}{\mathbb{E}[T|X = x, Z = \xi_k(x)] - \mathbb{E}[T|X = x, Z = \xi_{k-1}(x)]} \equiv \frac{\Upsilon_k(x)}{\rho_k(x)}$$

whenever $\rho_k(x) \equiv p(\xi_k(x), x) - p(\xi_{k-1}(x), x) \neq 0$. Assumption WM implies that each of these Wald estimands is a positively weighted average of what Angrist and Imbens (1995) call the average causal response, and thus is a positively weighted average of subgroup-specific treatment effects, and so weakly causal. Since $\phi_k(x) \geq 0$, whether these Wald estimands contribute to β_{rich} with positive or negative weight is determined by the sign of $\tilde{t}_k(x)$, which is the impact on T predicted by the first stage for a shift from $Z = \xi_{k-1}(X)$ to the values in $\Xi_k(x)$ that have larger propensity scores.[15] If $t_k(x)$ is non-negative, then the TSLS first stage is monotonicity-correct for at least one pair of instrument values, conditional on $X = x$.

We can isolate only the positively weighted contrasts by splitting β_{rich} into $\beta_{\text{rich}} = \beta^+_{\text{rich}} + \beta^-_{\text{rich}}$, where

$$\begin{aligned} \beta^+_{\text{rich}} &\equiv \mathbb{E}\left[\sum_{k=1}^{K} \Upsilon_k(X) \max\left\{0, \tilde{t}_k(X)\right\} \phi_k(X)\right] \\ \text{and} \qquad \beta^-_{\text{rich}} &\equiv \mathbb{E}\left[\sum_{k=1}^{K} \Upsilon_k(X) \min\left\{0, \tilde{t}_k(X)\right\} \phi_k(X)\right]. \end{aligned}$$

This extends decomposition (14) to

$$\beta_{\text{tsls}} = \overbrace{\beta^+_{\text{rich}}}^{\text{positively weighted}} + \underbrace{\beta^-_{\text{rich}}}_{\text{negatively weighted}} + \overbrace{(\beta_{\text{tsls}} - \beta_{\text{rich}})}^{\text{level-dependence}}.$$

[15] When defining $\xi_k(x)$, any ties in propensity scores are now broken based on the values of $\dot{t}(z, x)$. This breaks ties in favor of the empirical first stage and ensures that ties do not contribute to negative weights.

Only the first term, β^{+}_{rich}, reflects positively weighted treatment effects. The second term, β^{-}_{rich}, reflects negatively weighted treatment effects caused by a monotonicity-incorrect first stage, while the third term, $(\beta_{\text{tsls}} - \beta_{\text{rich}})$, captures level-dependence caused by insufficiently rich covariates. Each of the three terms in the decomposition are identified, enabling an empirical measurement of the makeup of the TSLS estimand for a given empirical specification.

While β^{+}_{rich} reflects positively weighted treatment effects, its weights do not necessarily sum to one. This has the undesirable implication that if treatment effects were indeed constant, β^{+}_{rich} would not equal that constant, making the interpretation of β^{+}_{rich} more difficult. To restore interpretation we also consider the rescaled version

$$\widetilde{\beta}^{+}_{\text{rich}} \equiv \beta^{+}_{\text{rich}} \times \mathbb{E}\left[\sum_{k=1}^{K} \rho_k(X) \max\left\{0, \tilde{t}_k(X)\right\} \phi_k(X)\right]^{-1},$$

and an analogous version $\widetilde{\beta}^{-}_{\text{rich}}$ of the negatively weighted component. The rescaled versions have weights that sum to one, so can be interpreted on the same scale as the subgroup-specific treatment effects.

6 Applications

In this section, we measure the extent to which commonly-employed TSLS specifications fail to produce weakly causal estimands by applying the decomposition in Proposition 11 to four empirical studies.

6.1 Gelbach (2002)

Gelbach (2002) estimates the impact of public school availability on maternal labor supply using a sample of single mothers whose youngest child was five years old in 1980. The outcome variable Y is maternal hours worked. The treatment variable T is an indicator for whether the mother's five-year-old was enrolled in public school. The instrument Z is the child's quarter of birth (QOB). The covariates X are race, residence in a central city, mother's age, state of birth, state of residence, number of own children and other household members by age. The sample size is 10,932.

Gelbach (2002, Table 2) provides evidence that QOB is correlated with some demographic variables, noting on pg. 309,

> *Since demographic variables are statistically associated with both QOB and labor supply, I control for them in the IV estimation.*

In his main results, Gelbach (2002, Table 6, column (4)) sets I to be indicators for each quarter of birth, and takes C to include everything in X as indicators except for age, which is specified as quadratic, number of own children, which is linear with different counts for ages 6–12, 13–17, and above 18, and number of other household members, which is linear with different counts for younger and older than 18. Gelbach's specification is not rich by construction; it has 111 variables, whereas a saturated specification has 10,397 variables, one for each covariate bin. Gelbach (2002, pg. 311) asserts that his estimates are nonparametric and reflect the compliers:

> *Subject to instrument validity, IV estimates nonparametrically identify the average effect of public schooling... [If] there is heterogeneity in the enrollment effects, only the effect of women observed enrolling their children is identified.*

Our results show that this claim is false. For Gelbach's estimates to reflect only compliers, the covariate specification must be rich, which requires making a *parametric* assumption about the conditional mean of the instruments given the covariates.

Column (2) of Table 3 reproduces Gelbach's (2002) main TSLS estimate and decomposes it using Proposition 11. The level-dependence portion is estimated to be almost as large as the original TSLS estimand. This means that if the employed specification had actually been rich, then the resulting estimate (β_{rich}) would have been nearly zero. Consistent with this finding, a RESET test provides strong evidence against the null hypothesis that the covariates are rich, with a p-value of .012.

Decomposing β_{rich} further into positively and negatively weighted components, we find that both components are small. This implies that even under constant treatment effects (Assumption CLE), the TSLS estimate in column (2) primarily reflects the levels of potential outcomes, rather than the differences (treatment effects). In order to restore a treatment effect interpretation, one could invoke Assumption LIN to impose a parametric assumption that the conditional means of potential outcomes are linear (Proposition 8).

Column (3) reports the TSLS estimate of the saturated, separable specification (SS). The estimate is closer to zero, consistent with the decomposition of the baseline specification, but is quite noisy. The reason is that saturating the covariates creates nearly as many covariate bins as there are observations. Only 789 observations lie in covariate bins with residual QOB variation—all others effectively get discarded in estimation due to perfect collinearity in the first stage. Columns (4) and (5) report TSLS and IJIVE estimates of the saturate and weight specification (SW), but these depend on an even smaller effective sample, and are also too noisy to learn much.

Table 3: Decomposition for Gelbach (2002, Table 6, column (4))

	(1)	(2)	(3)	(4)	(5)
	Original specification		Saturated covariate specifications		
	OLS	TSLS	TSLS (SS)	TSLS (SW)	IJIVE (SW)
Estimate (β_{tsls})	–1.009***	2.707***	0.738	–1.560	–1.410
	(0.381)	(0.881)	(11.290)	(6.391)	(6.607)
Level-dependence ($\beta_{\text{tsls}} - \beta_{\text{rich}}$)	—	2.666***	—	—	—
		[0.984, 4.429]			
Treatment effects (β_{rich})	—	0.041	0.738	–1.560	–1.410
		[–0.245, 0.324]	[–7.038, 9.282]		
Positively-weighted (β^{+}_{rich})	—	0.004	0.133	–1.560	–1.410
		[–0.266, 0.291]	[–7.458, 8.591]		
Negatively-weighted (β^{-}_{rich})	—	0.036	0.605	—	—
		[–0.027, 0.085]	[–0.816, 2.401]		
Causal effect, pos.-weighted ($\widetilde{\beta}^{+}_{\text{rich}}$)	—	0.067	0.119	—	—
		[–6.678, 7.486]	[–6.728, 7.637]		
Causal effect, neg.-weighted ($\widetilde{\beta}^{-}_{\text{rich}}$)	—	–5.190	–5.109	—	—
		[–19.504, 7.847]	[–19.361, 8.202]		
Ramsey RESET test p-val.					
($H_0:$ rich covariates)	—	0.012	—	—	—
Excluded variables	0	3	3	211	211
Included variables	111	111	10,397	10,397	10,397
Included variables, effective sample	111	111	349	186	186
Sample size	10,932	10,932	10,932	10,932	10,932
Effective sample size	10,932	10,932	789	440	440

Notes: The effective sample size is the number of observations that the estimator depends on after accounting for perfect collinearity in the first stage. Heteroskedasticity-robust standard errors are reported in parentheses for OLS, TSLS, and IJIVE estimates. Confidence intervals for the decomposition components are computed via nonparametric bootstrap based on 1000 bootstrap samples, with 95% confidence intervals shown in brackets. The RESET test is of the null hypothesis that $\mathbb{E}[\dot{Z}|X = x] = \mathbb{L}[\dot{Z}|C = c(x)]$ for all x. We implemented the RESET test using a nonparametric bootstrap estimate of the asymptotic variance matrix based on 1000 bootstrap samples. Stars *, **, and *** denote significance at levels .10, .05, and .01, respectively.

In Table 4, we consider a more parsimonious version of Gelbach's (2002) specification that includes only a quadratic in mother's age and state of residence fixed effects. Column (2) of Table 4 shows that the TSLS point estimate for this baseline specification is 2.752, which is nearly identical to Gelbach's original estimate of 2.707, and the standard error is only slightly larger. While the RESET test still provides some evidence that the specification does not have rich covariates, the impact on level-dependence is less severe, although still both economically and statistically significant.

Column (3) of Table 4 shows that there are still over 3,000 covariate bins in specification SS, even when only using mother's age and state of residence. However, the

Table 4: Decomposition of a parsimonious alternative to Gelbach (2002)

	(1)	(2)	(3)	(4)	(5)
	Baseline specification		Saturated covariate specifications		
	OLS	TSLS	TSLS (SS)	TSLS (SW)	IJIVE (SW)
Estimate (β_{tsls})	–2.987*** (0.397)	2.752*** (0.935)	2.474** (1.194)	–1.442** (0.689)	–0.340 (0.911)
Level-dependence ($\beta_{\text{tsls}} - \beta_{\text{rich}}$)	—	0.982** [0.024, 2.513]	—	—	—
Treatment effects (β_{rich})	—	1.770** [0.098, 2.889]	2.474** [0.230, 4.738]	–1.442	–0.340
Positively-weighted (β^{+}_{rich})	—	0.994 [–0.645, 2.005]	1.403 [–1.067, 3.352]	–1.442	–0.340
Negatively-weighted (β^{-}_{rich})	—	0.777*** [0.402, 1.319]	1.071*** [0.649, 2.211]	—	—
Causal effect, pos.-weighted ($\widetilde{\beta}^{+}_{\text{rich}}$)	—	1.275 [–0.976, 3.005]	1.289 [–0.955, 2.995]	—	—
Causal effect, neg.-weighted ($\widetilde{\beta}^{-}_{\text{rich}}$)	—	–12.128*** [–17.809, –5.491]	–12.069*** [–17.879, –5.529]	—	—
Ramsey RESET test p-val. ($H_0 :$ rich covariates)	—	0.124	—	—	—
Excluded variables	0	3	3	2,593	2,593
Included variables	52	52	3,177	3,177	3,177
Included variables, effective sample	52	52	1,808	1,329	1,329
Sample size	10,932	10,932	10,932	10,932	10,932
Effective sample size	10,932	10,932	9,341	7,975	7,975

Notes: Same notes as for Table 3.

effective sample is only about 10% smaller than the original sample. Consequently, the TSLS estimate for specification SS has a standard error roughly comparable to the more tightly parameterized estimate in column (2). The two point estimates are also fairly close, reflecting the lesser role of level-dependence in this more parsimonious specification. Under Assumption CLE, the SS estimate in column (3) is weakly causal.

Decomposing either column (2) or column (3) into positively and negatively weighted components shows that monotonicity-incorrectness is important here. The negatively weighted component is large and statistically significant for both estimates, implying that the separable first stage relationship used in specification SS fails to reproduce the sign of the nonparametric propensity score and thus is not monotonicity-correct. The negatively weighted component is positive because it reflects negatively weighted *negative* treatment effects for some groups. Reweighting the positive and negative components shows that a relatively small proportion receive negative weight, but that those that do have large negative treatment effects. As Gelbach (2002, pg. 308) notes,

Table 5: Sensitivity to covariate specification in Dube and Harish (2020)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Queen	1.011	0.511	0.681	0.984	1.220	0.262	1.190	0.400
	(0.523)	(0.231)	(0.355)	(0.519)	(0.640)	(0.170)	(0.639)	(0.211)
	[0.011]	[0.005]	[0.029]	[0.015]	[0.013]	[0.142]	[0.014]	[0.039]
Polity fixed effects		✓				✓		✓
Decade fixed effects			✓			✓		✓
Missing gender control				✓			✓	✓
Previous monarch controls					✓		✓	✓

Notes: Clustered standard errors are reported in parentheses. Brackets contain p-values for the clustered wild bootstrap procedure implemented by Dube and Harish (2020) with 1000 replications. Column (8) replicates Table 3, column (3) of Dube and Harish (2020). The sample size is 3,586.

the IV estimate reflects both the price and income effects of subsidized schooling, so its theoretical sign is ambiguous.

Column (4) of Table 4 reports specification SW, which is, by construction, a positively weighted average of complier treatment effects. The sign reverses from columns (2) and (3), providing further evidence of covariate-variation in the propensity score that is consistent with Assumption WM, but not with Assumption SM. However, the point estimate in column (4) is actually closer to the OLS estimate in (1), which could be the consequence of many instruments bias. Consistent with this explanation, the IJIVE estimate of specification SW in column (5) is considerably smaller in magnitude, and not statistically different from zero.

6.2 Dube and Harish (2020)

Dube and Harish (2020) estimate the effect of queenly rule on war using panel data on the polities of Europe covering the years 1480 to 1913. The outcome variable Y is a binary indicator for whether a polity-year observation was at war. The treatment variable T is a binary indicator for whether a queen ruled in that polity-year. Dube and Harish (2020) use two instruments Z; we focus on their preferred instrument, which is an indicator for whether the previous monarch had a legitimate firstborn male child. The covariates X in their main results (Dube and Harish, 2020, Table 3, column (3)) are polity and decade identifiers, whether the previous monarchs were corulers unrelated to one another, whether they had any legitimate children (with and without missing birth years), and whether the gender of the previous firstborn child is missing.

Dube and Harish (2020) justify most of their controls with concerns about exogeneity of the instrument. For example, they argue that controlling for whether the previous monarch had any legitimate children is necessary because the firstborn son instrument is mechanically zero whenever the previous monarch had no children (Dube and Harish, 2020, pp. 2601–2602). In Table 5 we show that without polity fixed effects their IV estimates are implausibly large, sometimes exceeding the logical value of 1, albeit with large standard errors. With both polity and decade fixed effects, but without the previous monarch controls, their estimates are close to half as large in magnitude. Covariates apparently matter substantially for the conclusions in this application.

Dube and Harish (2020, pg. 2605) explicitly invoke a LATE interpretation for their estimates:

> *If there are heterogeneous treatment effects, the IV estimate will be the LATE (Imbens and Angrist 1994). It will tell us the effect for the specific group of women who were eligible to rule and induced into ruling because of the presence of a firstborn female or sister among previous monarchs (i.e., the set of women who were compliers).*

However, their specifications do not include interactions between covariates, so they are not saturated, and thus are not necessarily rich. Their first stage specification also does not include any interactions between the instrument and the covariates, so may not be monotonicity-correct.

Table 6 reports our decomposition of Table 3, column (3) of Dube and Harish (2020). Column (2) replicates and decomposes their TSLS estimate. We find that almost all of the estimate is driven by level-dependence, as in Gelbach (2002). The RESET test overwhelmingly rejects the null hypothesis of rich covariates.

Column (3) of Table 6 reports specification SS, which has the opposite sign, but is imprecisely estimated. As in Gelbach (2002), this is because saturating the covariates creates a large number of bins. Few observations lie in covariate bins that have residual variation in the instrument, reducing the effective sample to only a fraction of the original sample. The implication is that the original TSLS specification is relying heavily on the assumed parametric form to extrapolate across covariate bins. The large level-dependence term and overwhelming rejection of the null of rich covariates in column (2) both suggest that this parametric form is working poorly here. Figure 2 provides additional evidence, showing that many of the fitted values of $\mathbb{L}[Z|C]$ lie outside of the interval $[0, 1]$ that $\mathbb{E}[Z|X]$ must lie in.

Columns (4) and (5) of Table 6 report TSLS and IJIVE estimates of specification SW. These depend on an even smaller subsample of 107 observations. The estimates are tighter than those for SS in column (3), but they are also close to the OLS estimate. This is suggestive of many instruments bias, which could be a problem here with 11

Table 6: Decomposition of Dube and Harish (2020, Table 3, column (3))

	(1)	(2)	(3)	(4)	(5)
	Original specification		Saturated covariate specifications		
	OLS	TSLS	TSLS (SS)	TSLS (SW)	IJIVE (SW)
Estimate (β_{tsls})	0.115***	0.400*	–0.509	0.145	0.148
	(0.035)	(0.211)	(0.523)	(0.099)	(0.101)
Level-dependence ($\beta_{\text{tsls}} - \beta_{\text{rich}}$)	—	0.447*	—	—	—
		[–0.007, 2.410]			
Treatment effects (β_{rich})	—	–0.047	–0.509	0.145	0.148
		[–0.288, 0.038]	[–7.021, 9.103]		
Positively-weighted (β^{+}_{rich})	—	–0.042	–0.451	0.145	0.148
		[–0.236, 0.043]	[–6.299, 10.477]		
Negatively-weighted (β^{-}_{rich})	—	–0.005	–0.058	—	—
		[–0.063, 0.022]	[–3.878, 0.423]		
Causal effect, pos.-weighted ($\widetilde{\beta}^{+}_{\text{rich}}$)	—	–0.294	–0.294	—	—
		[–3.355, 0.339]	[–1.427, 1.354]		
Causal effect, neg.-weighted ($\widetilde{\beta}^{-}_{\text{rich}}$)	—	0.109	0.109	—	—
		[–0.400, 0.667]	[–0.400, 0.667]		
Ramsey RESET test p-val.					
($H_0:$ rich covariates)	—	0.000	—	—	—
Excluded variables	0	1	1	11	11
Included variables	65	65	447	447	447
Included variables, effective sample	65	65	28	11	11
Sample size	3,586	3,586	3,586	3,586	3,586
Effective sample size	3,586	3,586	267	107	107

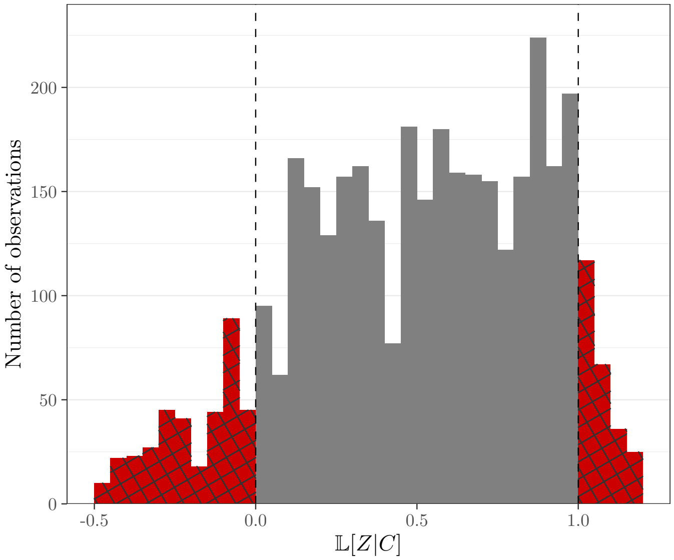
Notes: The effective sample size is the number of observations that the estimator depends on after accounting for perfect collinearity in the first stage. Clustered standard errors are reported in parentheses for OLS, TSLS, and IJIVE estimates. Confidence intervals for the decomposition components are computed via nonparametric block bootstrap based on 1000 bootstrap samples, with 95% confidence intervals shown in brackets. The RESET test is of the null hypothesis that $\mathbb{E}[Z|X = x] = \mathbb{L}[Z|C = c(x)]$ for all x. We implemented the RESET test using a cluster-robust estimate of the asymptotic variance matrix. Stars *, **, and *** denote significance at levels .10, .05, and .01, respectively.

excluded variables and only 107 observations. On the other hand, the IJIVE estimate is similar, suggesting that many instruments bias is not playing an important role. The Monte Carlo results in Appendix B show that IJIVE sometimes effectively corrects for many instruments bias and sometimes does not.

6.3 Card (1995)

Card (1995) uses a sample of 24-year-old men from the 1976 interview of the NLSY to estimate the returns to education. The outcome Y is log hourly wage. The treatment T is years of education. The instrument Z is a binary indicator for the presence of

Figure 2: Fitted values of $\mathbb{L}[Z|C]$ in Dube and Harish (2020)



Notes: A histogram of the fitted values from regressing $I = Z$ onto C for the TSLS specification in column (2) of Table 6.

an accredited four-year college in the local labor market when the respondent was 14 years old. In his main results, Card (1995, Table 3A, column (5)) specifies covariates C using a quadratic in years of potential experience, a race indicator for Black, geography indicators for living in the South and in an urban area, a set of indicators for region of residence in 1966, and an indicator for residence in an SMSA in 1966. All of these terms enter additively, so the covariate specification is not saturated, and thus not necessarily rich.

Column (2) of Table 7 reproduces Card's estimate of the returns to schooling and reports the results of our decomposition. The estimate in column (2) is weakly causal under the assumption of constant, linear effects, and a linearity assumption on the conditional mean of potential outcomes (Proposition 8). Without these assumptions, however, the specification must have rich covariates (Proposition 7). A RESET test overwhelmingly rejects the null hypothesis that the specification has rich covariates. In this case the impact on level-dependence is estimated to be fairly small, although the estimate is quite noisy. Figure 3 shows that most of the fitted values of $\mathbb{L}[Z|C]$ lie within the unit interval, as they should under correct specification.

Specification SS in column (3) has no level-dependence term by construction and produces an estimate that is modestly larger with a similar standard error. Under the assumption of constant, linear treatment effects the estimate in column (3) is weakly

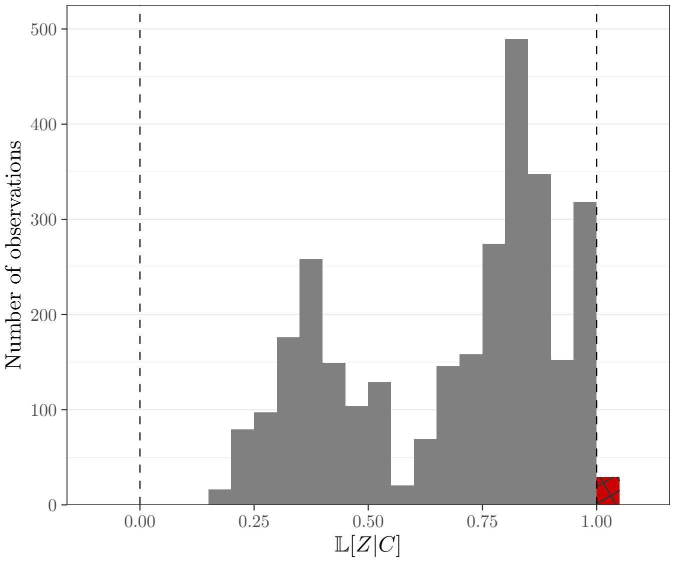
Table 7: Decomposition of Card (1995, Table 3A, column (5))

	(1)	(2)	(3)	(4)	(5)
	Original specification		Saturated covariate specifications		
	OLS	TSLS	TSLS (SS)	TSLS (SW)	IJIVE (SW)
Estimate (β_{tsls})	0.075***	0.132**	0.148**	0.072***	0.076***
	(0.004)	(0.054)	(0.064)	(0.013)	(0.021)
Level-dependence ($\beta_{\text{tsls}} - \beta_{\text{rich}}$)	—	0.020	—	—	—
		[–0.045, 0.137]			
Treatment effects (β_{rich})	—	0.112**	0.148**	0.072	0.076
		[0.007, 0.211]	[0.017, 0.327]		
Positively-weighted (β^{+}_{rich})	—	0.176***	0.234***	0.072	0.076
		[0.067, 0.294]	[0.104, 0.488]		
Negatively-weighted (β^{-}_{rich})	—	–0.064*	–0.085*	—	—
		[–0.148, 0.003]	[–0.272, 0.003]		
Causal effect, pos.-weighted ($\widetilde{\beta}^{+}_{\text{rich}}$)	—	0.109***	0.109***	—	—
		[0.049, 0.129]	[0.046, 0.133]		
Causal effect, neg.-weighted ($\widetilde{\beta}^{-}_{\text{rich}}$)	—	0.075*	0.075*	—	—
		[–0.005, 0.117]	[–0.002, 0.121]		
Ramsey RESET test p-val.					
($H_0:$ rich covariates)	—	0.000	—	—	—
Excluded variables	0	1	1	238	238
Included variables	15	15	819	819	819
Included variables, effective sample	15	15	264	238	238
Sample size	3,010	3,010	3,010	3,010	3,010
Effective sample size	3,010	3,010	1,864	1,780	1,780

Notes: The effective sample size is the number of observations that the estimator depends on after accounting for perfect collinearity in the first stage. Heteroskedasticity-robust standard errors are reported in parentheses for OLS, TSLS, and IJIVE estimates. Confidence intervals for the decomposition components are computed via nonparametric bootstrap based on 1000 bootstrap samples, with 95% confidence intervals shown in brackets. The RESET test is of the null hypothesis that $\mathbb{E}[Z|X = x] = \mathbb{L}[Z|C = c(x)]$ for all x. We implemented the RESET test using a heteroskedasticity-robust estimate of the asymptotic variance matrix. Stars *, **, and *** denote significance at levels .10, .05, and .01, respectively.

causal (Proposition 7). It is also weakly causal without constant, linear treatment effects if the monotonicity assumption is strengthened to Assumption SM, so that it operates in the same direction for all covariate groups (Proposition 10). Under weak monotonicity (Assumption WM), the estimate in column (3) could still reflect both positively and negatively weighted treatment effects. Our decomposition shows that both components are substantial in magnitude, with the overall TSLS estimate reflecting the sum of a large positively weighted component and a smaller negatively weighted component. Normalizing the weights of either component shows that they reflect treatment effect estimates that are closer to the OLS estimate than the original

Figure 3: Fitted values of $\mathbb{L}[Z|C]$ in Card (1995)



Notes: A histogram of the first-stage fitted values from regressing $I = Z$ onto C for the TSLS specification in column (2) of Table 7.

TSLS estimate.

Column (4) reports the TSLS estimate of specification SW, which will be weakly causal even under weak monotonicity (Assumption WM). The estimate is considerably smaller and more tightly estimated than either of the separable estimates in columns (3) and (4), and is even smaller than the OLS estimate in column (1). However, the large number of excluded variables raises concerns about many instruments bias. Applying IJIVE to the same specification increases the estimate slightly to be almost identical to the OLS estimate. This suggests that if there is many instruments bias, IJIVE fails to correct for it, which is consistent with our simulation results for this sample size and number of covariate bins (see Appendix B).

6.4 Angrist and Krueger (1991)

Angrist and Krueger (1991) estimate the returns to education using a sample of men from the 1980 Census who were born in the United States between 1930 and 1939. The outcome variable Y is log weekly earnings. The treatment variable T is years of education. The instrument Z is quarter of birth (QOB). Covariates X are year of birth, region of residence, state of birth, race, marital status, and residing in an urban area.

We decompose the estimate in Table V, column (6) of Angrist and Krueger (1991).

Table 8: Decomposition of Angrist and Krueger (1991, Table V, column (6))

	(1)	(2)	(3)	(4)	(5)
	Original specification		Saturated covariate specifications		
	OLS	TSLS	TSLS (SS)	TSLS (SW)	IJIVE (SW)
Estimate (β_{tsls})	0.063***	0.081***	0.078***	0.061***	0.063**
	(0.000)	(0.016)	(0.017)	(0.005)	(0.029)
Level-dependence ($\beta_{\text{tsls}} - \beta_{\text{rich}}$)	—	0.002	—	—	—
		[–0.001, 0.005]			
Treatment effects (β_{rich})	—	0.078***	0.078***	0.061	0.063
		[0.045, 0.108]	[0.046, 0.109]		
Positively-weighted (β^{+}_{rich})	—	0.101***	0.100***	0.061	0.063
		[0.076, 0.134]	[0.077, 0.133]		
Negatively-weighted (β^{-}_{rich})	—	–0.022***	–0.022***	—	—
		[–0.051, –0.012]	[–0.048, –0.011]		
Causal effect, pos.-weighted ($\widetilde{\beta}^{+}_{\text{rich}}$)	—	0.075***	0.074***	—	—
		[0.051, 0.087]	[0.052, 0.088]		
Causal effect, neg.-weighted ($\widetilde{\beta}^{-}_{\text{rich}}$)	—	0.063***	0.063***	—	—
		[0.024, 0.090]	[0.022, 0.090]		
Ramsey RESET test p-val. ($H_0 :$ rich covariates)	—	0.837	—	—	—
Excluded variables	0	30	30	1,900	1,900
Included variables	21	21	695	695	695
Included variables, effective sample	21	21	661	659	659
Sample size	329,509	329,509	329,509	329,509	329,509
Effective sample size	329,509	329,509	329,468	329,463	329,463

Notes: The effective sample size is the number of observations that the estimator depends on after accounting for perfect collinearity in the first stage. Heteroskedasticity-robust standard errors are reported in parentheses for OLS, TSLS, and IJIVE estimates. Confidence intervals for the decomposition components are computed via nonparametric bootstrap based on 1000 bootstrap samples, with 95% confidence intervals shown in brackets. The RESET test is of the null hypothesis that $\mathbb{E}[\dot{Z}|X = x] = \mathbb{L}[\dot{Z}|C = c(x)]$ for all x. We implemented the RESET test using a nonparametric bootstrap estimate of the asymptotic variance matrix based on 1000 bootstrap samples. Stars *, **, and *** denote significance at levels .10, .05, and .01, respectively.

For this specification the excluded variables I are QOB interacted with indicators for year of birth. The covariate specification C is a full non-interacted set of indicators for year of birth, region of residence, and state of birth, as well as non-interacted indicators for race, marital status, and residence in an SMSA. This specification is not necessarily rich in general. However, if QOB is independent of the covariates—which seems plausible, at least on first consideration—then *any* specification C that includes a constant term is rich.[16]

[16] Angrist and Krueger (1991) cite Lam and Miron (1987, 1991) as providing evidence that parents' socioeconomic status and other characteristics are unrelated to season of birth. Buckles and Hungerman (2013)

Table 8 reports the decomposition results. The RESET test fails to reject the null of rich covariates, and we find essentially no level-dependence. This implies that the estimate is weakly causal under the assumption of constant, linear treatment effects (Assumption CLE), and leads the TSLS estimate of specification SS to be similar. If we allow for heterogeneous treatment effects, however, the decomposition shows that roughly 25% of the original estimate is due to negatively weighted positive treatment effects. The positively and negatively weighted treatment effect components are smaller than the overall TSLS estimate, with the negatively weighted component approximately the same as the OLS estimate.

The negatively weighted component is mechanically removed by the introduction of covariate-instrument interactions in specification (SW), reported in column (4). The estimate is tightly estimated and remarkably close to the OLS estimate in column (1), raising concerns of many instruments bias. The IJIVE estimate in column (5) is similar but with a much larger standard error, suggesting that if there is indeed many instruments bias, it fails to correct for it.

7 Conclusion

In discussing the LATE interpretation of (linear) IV estimates, Angrist and Krueger (1999, pg. 1326) conjectured:

> *That is, IV estimates in models with covariates can be thought of as producing a weighted average of covariate-specific Wald estimates as long as the model for covariates is saturated . . . . In other cases it seems reasonable to assume that some sort of approximate weighted average is being generated, but we are unaware of a precise causal interpretation that fits all cases.*

In this paper we have provided the precise—sufficient and necessary—causal interpretation that fits (nearly) all cases. The interpretation shows that TSLS estimates with covariates *cannot* be interpreted as "weakly causal," and thus not as a positively weighted averages of LATEs, at least not without additional parametric assumptions. We developed a decomposition that measures the extent to which a TSLS estimand fails to be weakly causal and applied it to four empirical applications, finding evidence in each that in practice, the LATE interpretation of TSLS does not hold.

Our findings show that TSLS only has a causal interpretation under parametric assumptions or when using close-to-nonparametric specifications, such as specifications SS or SW. In this regard TSLS has no advantage over more explicit parametric (Im-

find evidence of a relationship between socioeconomic status and season of birth in births after 1943.

bens and Rubin, 1997; Hirano et al., 2000; Yau and Little, 2001; Słoczyński et al., 2022), semiparametric (Abadie, 2003; Tan, 2006, 2010; Hong and Nekipelov, 2010), and nonparametric (Frölich, 2007; Ogburn et al., 2015; Chernozhukov et al., 2018; Athey et al., 2019; Heiler, 2021; Sun and Tan, 2021; Singh and Sun, 2022) approaches to estimating LATEs. Researchers may want to consider adopting one of these more principled approaches. Alternatively, the implicit parametric assumptions invoked in non-saturated TSLS specifications should be defended, both by stating them explicitly, and by evaluating their adequacy through specification testing.

It is important to reemphasize that the criterion of "weakly causal" used throughout the analysis is an extremely weak one. Being weakly causal may be necessary for a quantity to represent an interesting causal effect, but it is not sufficient. For example, Słoczyński (2022) argues that even the saturated specification SS produces an estimand that can be difficult to interpret and may be quite different from the unconditional LATEs considered in the original Imbens and Angrist (1994) analysis.

These interpretation difficulties were already reason to recommend alternative IV methods designed to estimate quantities, such as the average treatment on the treated, that are not only weakly causal but also have clear counterfactual interpretations. Such methods rely on explicitly stated parametric assumptions (Heckman, 1976; Heckman et al., 2003) or are semiparametric (Carneiro et al., 2011; Brinch et al., 2017; Mogstad et al., 2018) or nonparametric (e.g. Heckman and Vytlacil, 1999; Manski and Pepper, 2000). By showing that common interpretations of TSLS also rely on either parametric assumptions or nonparametric implementations, our findings provide another reason to pursue such approaches.

A Proofs

Proof of Proposition 1. The expression for β_{iv} is a special case of Proposition 2 with $\epsilon = 1$.

If $\mathbb{E}[T\tilde{Z}] > 0$, then because $\mathbb{E}[Z|X] \in [0,1]$ for binary Z, the sign of $\omega(\textsc{cp}, X)$ depends on the sign of $1 - \mathbb{L}[Z|X]$, which is negative if and only if $\mathbb{L}[Z|X] > 1$. The sign of $\omega(\textsc{at}, X)$ varies with X according to the sign of $\mathbb{E}[\tilde{Z}|X]$. Because X contains a constant, $\mathbb{E}[\mathbb{E}[\tilde{Z}|X]] = \mathbb{E}[\tilde{Z}] = 0$, and thus $\mathbb{E}[\tilde{Z}|X]$ is either zero with probability 1, or else it has positive probability of taking both positive and negative values. In the latter case, the sign of $\omega(\textsc{at}, X)$ is negative for some values of X regardless of whether $\mathbb{E}[T\tilde{Z}]$ is positive or negative. *Q.E.D.*

Proof of Proposition 2. The numerator of β_{iv} can be written as

$$\mathbb{E}[Y\tilde{Z}] = \mathbb{E}\left[\mathbb{E}\left[Y\tilde{Z}|X\right]\right] = \mathbb{E}\left[\mathbb{C}[Y,Z|X]\right] + \mathbb{E}\left[\mathbb{E}[Y|X]\,\mathbb{E}[\tilde{Z}|X]\right]. \tag{15}$$

The same argument as in Imbens and Angrist (1994) applied conditional-on-covariates yields

$$\mathbb{C}[Y,Z|X] = \mu(\textsc{cp},X)\,\mathbb{C}[T,Z|X] = \mu(\textsc{cp},X)\,\mathbb{P}[G=\textsc{cp}|X]\,\mathbb{E}[Z|X](1-\mathbb{E}[Z|X]). \tag{16}$$

As for the second term of (15),

$$\begin{aligned}\mathbb{E}\left[Y|X\right] &= \mathbb{E}\left[Y|G=\textsc{at},X\right]\mathbb{P}\left[G=\textsc{at}|X\right] + \mathbb{E}\left[Y|G=\textsc{nt},X\right]\mathbb{P}\left[G=\textsc{nt}|X\right]\\ &\quad + \mathbb{E}\left[Y|G=\textsc{cp},X\right]\mathbb{P}\left[G=\textsc{cp}|X\right]\\ &= \mathbb{E}\left[Y(1)|G=\textsc{at},X\right]\mathbb{P}\left[G=\textsc{at}|X\right] + \mathbb{E}\left[Y(0)|G=\textsc{nt},X\right]\mathbb{P}\left[G=\textsc{nt}|X\right]\\ &\quad + \mathbb{E}\left[(1-Z)Y(0)+ZY(1)|G=\textsc{cp},X\right]\mathbb{P}\left[G=\textsc{cp}|X\right].\end{aligned} \tag{17}$$

Adding and subtracting $\mathbb{E}[Y(0)|G=\textsc{at},X]\,\mathbb{P}[G=\textsc{at}|X]$ gives

$$\mathbb{E}\left[Y|X\right] = \Delta(\textsc{cp},X)\,\mathbb{P}[G=\textsc{cp}|X]\,\mathbb{E}[Z|X] + \Delta(\textsc{at},X)\,\mathbb{P}[G=\textsc{at}|X] + \eta_0'X \tag{18}$$

due to both the exogeneity of Z and the linearity assumption on $\mathbb{E}[Y(0)|X=x]$. Alternatively, adding and subtracting $\mathbb{E}[Y(1)|G=\textsc{nt},X]\,\mathbb{P}[G=\textsc{nt}|X]$ to (17) gives

$$\mathbb{E}\left[Y|X\right] = \Delta(\textsc{cp},X)\,\mathbb{P}[G=\textsc{cp}|X](\mathbb{E}[Z|X]-1) - \Delta(\textsc{nt},X)\,\mathbb{P}[G=\textsc{nt}|X] + \eta_1'X. \tag{19}$$

So multiplying (18) by ϵ and summing it with (19) multiplied by $1-\epsilon$ gives

$$\begin{aligned}\mathbb{E}[Y|X] = \Delta(\textsc{cp}, X)\,\mathbb{P}[G = \textsc{cp}|X]\,(\mathbb{E}[Z|X] + \epsilon - 1) + \Delta(\textsc{at}, X)\epsilon\,\mathbb{P}[G = \textsc{at}|X] \\ + \Delta(\textsc{nt}, X)(\epsilon - 1)\,\mathbb{P}[G = \textsc{nt}|X] + \epsilon\eta_0' X + (1-\epsilon)\eta_1' X.\end{aligned}$$

Because X and $\tilde{Z}$ are orthogonal,

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}[Y|X]\,\mathbb{E}[\tilde{Z}|X]\right] = \mathbb{E}\Big[&\Delta(\textsc{cp}, X)\,\mathbb{P}[G = \textsc{cp}|X]\,(\mathbb{E}[Z|X] + \epsilon - 1)\,\mathbb{E}[\tilde{Z}|X] \\ &+ \Delta(\textsc{at}, X)\epsilon\,\mathbb{P}[G = \textsc{at}|X]\,\mathbb{E}[\tilde{Z}|X] \\ &+ \Delta(\textsc{nt}, X)(\epsilon - 1)\,\mathbb{P}[G = \textsc{nt}|X]\,\mathbb{E}[\tilde{Z}|X]\Big]. \qquad (20)\end{aligned}$$

Summing (16) and (20), and noting that

$$\begin{aligned}&\mathbb{E}[Z|X](1 - \mathbb{E}[Z|X]) + (\mathbb{E}[Z|X] + \epsilon - 1)\,\mathbb{E}[\tilde{Z}|X] \\ &\quad = \left(\mathbb{E}[Z|X] - \mathbb{E}[\tilde{Z}|X]\right)(1 - \mathbb{E}[Z|X]) + \epsilon\,\mathbb{E}[\tilde{Z}|X] \\ &\quad = \mathbb{L}[Z|X](1 - \mathbb{E}[Z|X]) + \epsilon\,\mathbb{E}[\tilde{Z}|X]\end{aligned}$$

yields a weighting expression with weights proportional to the claimed expression but missing a common multiple of $\mathbb{E}[\tilde{Z}T]^{-1}$, which comes from the denominator of β_{iv}. *Q.E.D.*

Proof of Proposition 3. Note that T is only stochastic due to Z after conditioning on X and G, as a direct consequence of the definition of G. Thus, Assumption EX implies that T and $Y(t)$ are independent conditional on X and G. We use this observation to write

$$\begin{aligned}\beta &= \sum_{g,x} \mathbb{E}\left[b(T, x, Z)Y \middle| G = g, X = x\right]\mathbb{P}[G = g, X = x] \\ &= \sum_{g,x,j} \mathbb{E}\left[\mathbb{1}[T = t_j]b(t_j, x, Z)Y(t_j) \middle| G = g, X = x\right]\mathbb{P}[G = g, X = x] \\ &= \sum_{g,x,j} \mu_j(g, x)\,\mathbb{E}\left[\mathbb{1}[T = t_j]b(t_j, x, Z) \middle| G = g, X = x\right]\mathbb{P}[G = g, X = x] \\ &\equiv \sum_{g,x,j} \mu_j(g, x)\psi_j(g, x),\end{aligned}$$

where all summations are taken over $g \in \mathcal{G}, x \in \mathcal{X}, j \in \{0, 1, \ldots, J\}$, and

$$\psi_j(g, x) \equiv \mathbb{E}\left[\mathbb{1}[T = t_j]b(t_j, x, Z) \middle| G = g, X = x\right]\mathbb{P}[G = g, X = x].$$

Notice that $\omega_j(g,x) = \sum_{k=j}^{J} \psi_k(g,x)$, so that (6) follows from Lemma 1.

Q.E.D.

Lemma 1. For any constants $\{a_j, c_j\}_{j=0}^{J}$,

$$\sum_{j=0}^{J} a_j c_j = a_0 \tilde{c}_0 + \sum_{j=1}^{J} (a_j - a_{j-1}) \tilde{c}_j,$$

where $\tilde{c}_j \equiv \sum_{k=j}^{J} c_k$.

Proof of Lemma 1. Since $c_j = \tilde{c}_j - \tilde{c}_{j+1}$,

$$\begin{aligned}\sum_{j=0}^{J} a_j c_j &= \sum_{j=0}^{J} a_j \left(\tilde{c}_j - \tilde{c}_{j+1}\right) \\ &= a_0 \tilde{c}_0 + \sum_{j=1}^{J} a_j \tilde{c}_j + \sum_{j=0}^{J-1} a_j \tilde{c}_{j+1} = a_0 \tilde{c}_0 + \sum_{j=1}^{J} (a_j - a_{j-1}) \tilde{c}_j,\end{aligned}$$

where the final equality used a change of variables in the second summand from j to $j+1$. *Q.E.D.*

Proof of Proposition 4. If $\omega_j(g,x) \geq 0$ and $\omega_0(g,x) = 0$ for all g and x, then it follows immediately from (6) that β satisfies Definition WC.

We will prove the converse by contraposition. That is, we will show that if either the non-negative weights or level irrelevance condition is not satisfied, then there exists a $\mu \in \mathcal{M} = \mathcal{M}_{\Box}$ such that $\mu_j(g,x) - \mu_{j-1}(g,x)$ has the same sign for every $j \geq 1$, and all g and x, and that this common sign is different than the sign of β. Thus, if the weights do not satisfy both the non-negative and level irrelevance conditions, then β is not weakly causal. Or, by contraposition, if β is weakly causal, then the weights satisfy both conditions.

First, suppose that the level irrelevance condition does not hold, but that the non-negative weights condition may or may not hold. Then there exists a $(g^\star, x^\star)$ such that $\omega_0(g^\star, x^\star) \neq 0$. Set

$$\mu_j(g,x) = \begin{cases} \bar{\mu}, & \text{if } (g,x) \neq (g^\star, x^\star) \\ \mu^\star, & \text{if } (g,x) = (g^\star, x^\star) \text{ and } j < j^\star \ , \\ \mu^\star + \Delta^\star, & \text{if } (g,x) = (g^\star, x^\star) \text{ and } j \geq j^\star \end{cases} \tag{21}$$

where $\bar{\mu}, \mu^\star \in (\underline{y}, \overline{y})$ and $\Delta^\star \in (\underline{y} - \overline{y}, \overline{y} - \underline{y})$ are numbers we will choose, and $j^\star \geq 1$ can be chosen arbitrarily. Then $\mu_j(g,x) - \mu_{j-1}(g,x)$ is zero for all $(g,x) \neq (g^\star, x^\star)$, while for $(g,x) = (g^\star, x^\star)$ it is $\Delta^\star$ when $j = j^\star$ and zero otherwise. In particular, the sign of $\mu_j(g,x) - \mu_{j-1}(g,x)$ is the sign of $\Delta^\star$ for all $j \geq 1$ and all (g,x), regardless of the values of $\bar{\mu}$ and $\mu^\star$. If μ is specified as in (21), then (6) becomes

$$\beta = \underbrace{\left[\left(\sum_{(g,x)\neq(g^\star,x^\star)} \omega_0(g,x)\right)\bar{\mu} + \omega_0(g^\star,x^\star)\mu^\star\right]}_{\text{define this to be } \beta_0(\bar{\mu},\mu^\star) \text{ for shorthand}} + \omega_{j^\star}(g^\star,x^\star)\Delta^\star. \tag{22}$$

Since $\omega_0(g^\star,x^\star) \neq 0$, there exist values of $\bar{\mu}$ and $\mu^\star$ such that $\beta_0(\bar{\mu},\mu^\star) \neq 0$. Choose any such $\bar{\mu}, \mu^\star \in (\underline{y}, \overline{y})$. If $\beta_0(\bar{\mu},\mu^\star) > 0$, then choose a $\Delta^\star < 0$ that is sufficiently small in magnitude so that $\omega_{j^\star}(g^\star,x^\star)\Delta^\star > -\beta_0(\bar{\mu},\mu^\star)$ and $\mu^\star + \Delta \in [\underline{y}, \overline{y}]$. Then from (22) we have $\beta = \beta_0(\bar{\mu},\mu^\star) + \omega_{j^\star}(g^\star,x^\star)\Delta^\star > 0$, so that these choices of $\bar{\mu}$, $\mu^\star$, and $\Delta^\star$ produce a $\mu \in \mathcal{M}_\Box$ that violates the second condition of Definition WC. Similarly, if $\beta_0(\bar{\mu},\mu^\star) < 0$, then choose $\Delta^\star > 0$ to be sufficiently small to ensure that $\omega_{j^\star}(g^\star,x^\star)\Delta^\star < -\beta_0(\bar{\mu},\mu^\star)$, so that $\beta < 0$, contradicting the first condition of Definition WC.

On the other hand, suppose that the level irrelevance condition holds. Then, by hypothesis, the non-negative weights condition does not hold, so there exists a $j^\star, g^\star$, and $x^\star$ such that $\omega_{j^\star}(g^\star,x^\star) < 0$. Use the same construction as in (21) with these new values of $j^\star, g^\star$, and $x^\star$, where $j^\star$ is no longer arbitrary. Because the level irrelevance condition holds, (22) reduces to

$$\beta = \omega_{j^\star}(g^\star,x^\star)\Delta^\star.$$

Selecting any $\Delta^\star > 0$ produces $\beta < 0$, thus providing the existence of a $\mu \in \mathcal{M}_\Box$ that violates the first condition of Definition WC. *Q.E.D.*

Proof of Proposition 5. If Assumption CLE is satisfied, so that $\mathcal{M} = \mathcal{M}_{\text{CLE}}$, then

$$\mu_j(g,x) - \mu_{j-1}(g,x) = \Delta(t_j - t_{j-1})$$

for every $j \geq 1$, g, and x. Substituting into (6) yields

$$\beta = \sum_{g,x} \omega_0(g,x)\mu_0(g,x) + \left(\sum_{g,x}\sum_{j=1}^{J}(t_j - t_{j-1})\omega_j(g,x)\right)\Delta \tag{23}$$

If $\Delta \geq 0$, so that all treatment effects are positive, and if the level irrelevance and aggregated non-negative weights conditions hold, then (23) shows that $\beta \geq 0$ as well, and thus β satisfies Definition WC.

To prove the converse, we follow the same strategy as in the proof of Proposition 4 of showing that if the weights fail either the aggregated non-negativity or level irrelevance conditions, then there always exists a $\mu \in \mathcal{M} = \mathcal{M}_{\text{CLE}}$ that contradicts one of the conditions in Definition WC (i.e. (4)).

First, suppose that the level irrelevance condition does not hold, so that there exists a $(g^\star, x^\star)$ such that $\omega_0(g^\star, x^\star) \neq 0$. Set

$$\mu_j(g,x) = \begin{cases} \bar{\mu}, & \text{if } (g,x) \neq (g^\star, x^\star) \text{ and } j = 0 \\ \bar{\mu} + \Delta(t_j - t_0), & \text{if } (g,x) \neq (g^\star, x^\star) \text{ and } j > 0 \\ \mu^\star, & \text{if } (g,x) = (g^\star, x^\star) \text{ and } j = 0 \\ \mu^\star + \Delta(t_j - t_0), & \text{if } (g,x) = (g^\star, x^\star) \text{ and } j > 0 \end{cases},$$

for $\bar{\mu}, \mu^\star \in (\underline{y}, \overline{y})$ and $\Delta \in [\underline{y} - \overline{y}, \overline{y} - \underline{y}]$. Notice that $\mu \in \mathcal{M}_{\text{CLE}}$ if Δ is sufficiently small in magnitude. From (23),

$$\beta = \beta_0(\bar{\mu}, \mu^\star) + \bar{\omega}\Delta, \tag{24}$$

where $\beta_0(\bar{\mu}, \mu^\star)$ is the same quantity defined in (22) and

$$\bar{\omega} \equiv \sum_{g,x} \sum_{j=1}^{J} (t_j - t_{j-1}) \omega_j(g,x).$$

Since $\omega_0(g^\star, x^\star) \neq 0$, there exist values of $\bar{\mu}$ and $\mu^\star$ such that $\beta_0(\bar{\mu}, \mu^\star) \neq 0$. If $\beta_0(\bar{\mu}, \mu^\star) > 0$, then choose $\Delta < 0$ to be sufficiently small in magnitude to make $\bar{\omega}\Delta > -\beta_0(\bar{\mu}, \mu^\star)$, and thus $\beta > 0$, violating the second condition of Definition WC. If $\beta_0(\bar{\mu}, \mu^\star) < 0$, then choosing a small $\Delta > 0$ violates the first condition of Definition WC.

If the level-irrelevance condition does not hold then, by hypothesis, the non-negative weights condition is not satisfied, so that $\bar{\omega} < 0$. Since $\beta = \bar{\omega}\Delta$ under level-irrelevance, taking any $\Delta > 0$ creates a violation of the first condition of Definition WC. *Q.E.D.*

Proof of Proposition 6. The well-known two stage interpretation of α_{tsls} is

$$\alpha_{\text{tsls}} = \mathbb{E}[\dot{S}\dot{S}']^{-1}\,\mathbb{E}[\dot{S}Y], \tag{25}$$

where $\dot{S} \equiv \mathbb{E}[SF']\,\mathbb{E}[FF']^{-1}F$. Since C is a subvector of both S and F, $\dot{S} = [\dot{T}', C]'$, where $\dot{T}$ is the population fitted value from the first stage regression of T on I and C.

This fitted value can be written as

$$\dot{T} = \gamma' I + \lambda' C \equiv \dot{Z} + \lambda' C, \tag{26}$$

where γ and λ are population regression coefficients. Applying the Frisch-Waugh-Lovell Theorem to the second step regression (with full vector of coefficients (25)), the component of α_{tsls} corresponding to the coefficient on $\dot{T}$ can be written as

$$\beta_{\text{tsls}} = \mathbb{E}[RY] / \mathbb{E}[R^2],$$

where $R \equiv \dot{T} - \mathbb{L}[\dot{T}|C]$ are the residuals from projecting the population fitted treatment variable, $\dot{T}$, onto the covariates, C. Using (26), these residuals can be written more simply as

$$R \equiv \dot{T} - \mathbb{L}[\dot{T}|C] = \left(\dot{Z} + \lambda' C\right) - \mathbb{L}\left[\dot{Z} + \lambda' C | C\right] = \dot{Z} - \mathbb{L}[\dot{Z}|C] \equiv \tilde{Z}.$$

This shows that $\beta_{\text{tsls}} = \mathbb{E}[\tilde{Z}Y] / \mathbb{E}[\tilde{Z}^2]$. Since $\tilde{Z}$ is a residual from a projection onto C, we can also use (26) to write

$$\mathbb{E}[\tilde{Z}^2] = \mathbb{E}[\tilde{Z}\dot{Z}] = \mathbb{E}[\tilde{Z}(\dot{T} - \lambda' C)] = \mathbb{E}[\tilde{Z}T] - \mathbb{E}[\tilde{Z}(T - \dot{T})] = \mathbb{E}[\tilde{Z}T],$$

where the final equality follows because $\tilde{Z}$ is a linear function of I and C, and thus orthogonal to the first stage residuals, $T - \dot{T}$. *Q.E.D.*

Proof of Proposition 7. We evaluate the sufficient and necessary conditions in Proposition 5 using the expressions for $\omega_j(g, x)$ given in (9).

First, consider the aggregated non-negative weights condition in Proposition 5. Then since $\omega_{J+1}(g, x) = 0$,

$$\begin{aligned} \sum_{g,x} \sum_{j=1}^{J} (t_j - t_{j-1}) \omega_j(g, x) &= \sum_{g,x} \left(\sum_{j=1}^{J} t_j \omega_j(g, x) - \sum_{j=0}^{J-1} t_j \omega_{j+1}(g, x) \right) \\ &= \sum_{g,x} \sum_{j=0}^{J} t_j (\omega_j(g, x) - \omega_{j+1}(g, x)) - t_0 \sum_{g,x} \omega_0(g, x). \end{aligned}$$

The second term is zero when $\omega_j(g, x)$ is given by (9) because

$$\sum_{g,x} \omega_0(g, x) = \mathbb{E}[\tilde{Z}T]^{-1} \mathbb{E}[\tilde{Z}] = 0,$$

due to the assumption that C contains a constant regressor. The first term satisfies

$$\sum_{g,x}\sum_{j=0}^{J} t_j(\omega_j(g,x) - \omega_{j+1}(g,x)) = \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}\left[\left(\sum_{j=0}^{J} t_j \mathbb{1}[T = t_j]\right)\tilde{Z}\right] = 1,$$

so is also non-negative, as required.

Second, consider the level irrelevance condition, which given (9) can be written as

$$\omega_0(g,x) = \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}[\tilde{Z}|G = g, X = x]\,\mathbb{P}[G = g, X = x] = 0 \tag{27}$$

for all g and x. Assumption EX implies that $\tilde{Z} \equiv \dot{Z} - \mathbb{L}[\dot{Z}|C]$ is independent of G given X, so

$$\mathbb{E}[\tilde{Z}|G = g, X = x] = \mathbb{E}[\tilde{Z}|X = x] = \mathbb{E}[\dot{Z}|X = x] - \mathbb{L}[\dot{Z}|C = c(x)].$$

For every x there exists a $g \in \mathcal{G}$ such that $\mathbb{P}[G = g, X = x] > 0$, because G exhaustively partitions possible choice types. Thus, (27) can hold for every g and x if and only if

$$\mathbb{E}[\dot{Z}|X = x] = \mathbb{L}[\dot{Z}|C = c(x)]$$

for every x, that is, if and only if the TSLS specification has rich covariates.

We have shown that the aggregated non-negative weights condition is satisfied whether or not the TSLS specification has rich covariates, and that the level irrelevance condition is satisfied if and only if the TSLS specification has rich covariates. The claim now follows from Proposition 5.

Q.E.D.

Proof of Proposition 8. Under Assumption CLE,

$$Y = \sum_{j=0}^{J} Y(t_j)\mathbb{1}[T = t_j] = Y(t_0) + \sum_{j=1}^{J}(Y(t_j) - Y(t_0))\,\mathbb{1}[T = t_j] = Y(t_0) + \Delta(T - t_0),$$

so that

$$\beta_{\text{tsls}} = \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}[\tilde{Z}(Y(t_0) + (\Delta T - t_0)] = \Delta + \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}[\tilde{Z}Y(t_0)].$$

Given Assumption CLE, Assumption LIN also implies that

$$\mathbb{E}[Y(t_0)|X = x] = \mathbb{E}[Y(t_j) - Y(t_0)|X = x] + \mathbb{E}[Y(t_j)|X = x] = \Delta(t_j - t_0) + \eta' c(x),$$

so that $\mathbb{E}[Y(t_0)|X = x] = \eta_0' c(x)$, where η_0 is the same as η but has $\Delta(t_j - t_0)$ added to the coefficient on the constant regressor. Because $\tilde{Z}$ is orthogonal to C,

$$\mathbb{E}[\tilde{Z}Y(t_0)] = \mathbb{E}[\tilde{Z}\,\mathbb{E}[Y(t_0)|X]] = \mathbb{E}[\tilde{Z}C]'\eta_0 = 0,$$

so that $\beta_{\text{tsls}} = \Delta$, as claimed. *Q.E.D.*

Proof of Proposition 9. We evaluate the sufficient and necessary conditions given in Proposition 4 using the expressions for $\omega_j(g, x)$ given in (9). One of the conditions in Proposition 4, level irrelevance, was shown to be satisfied in the proof of Proposition 7 if the TSLS specification has rich covariates and C contains a constant, both of which are maintained assumptions here. Thus, we turn our focus to the other requirement of non-negative weights, which is that $\omega_j(g, x) \geq 0$ for all g, x, and $j \geq 1$.

We use the observation that an individual's choice group G completely determines what treatment value they would choose as a function of the instrument realization, Z. Let $\mathcal{Z}_j(g)$ denote the set of instrument values for which individuals in choice group g would choose a treatment value t_j or larger. Then

$$\begin{aligned}
\omega_j(g, x) &= \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}\left[\tilde{Z}\mathbb{1}[T \geq t_j]\middle| G = g, X = x\right]\mathbb{P}[G = g, X = x] \\
&= \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}\left[\tilde{Z}\mathbb{1}[Z \in \mathcal{Z}_j(g)]\middle| G = g, X = x\right]\mathbb{P}[G = g, X = x] \\
&= \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{E}\left[\tilde{Z}\mathbb{1}[Z \in \mathcal{Z}_j(g)]\middle| X = x\right]\mathbb{P}[G = g, X = x] \\
&= \mathbb{E}[\tilde{Z}T]^{-1}\,\mathbb{C}\left[\dot{Z}, \mathbb{1}[Z \in \mathcal{Z}_j(g)]\middle| X = x\right]\mathbb{P}[G = g, X = x], \qquad (28)
\end{aligned}$$

where the third equality used Assumption EX and the fourth used the assumption that the TSLS specification has rich covariates, so that $\tilde{Z}$ has mean zero given $X = x$.

In the proof of Proposition 6 we showed that $\mathbb{E}[\tilde{Z}T] = \mathbb{E}[\tilde{Z}^2] > 0$, so (28) implies that $\omega_j(g, x) \geq 0$ if and only if $\dot{Z}$ and $\mathbb{1}[Z \in \mathcal{Z}_j(g)]$ are positively correlated, conditional on $X = x$, for every (g, x) pair such that $\mathbb{P}[G = g, X = x] > 0$. Because

$$\begin{aligned}
&\mathbb{C}\left[\dot{Z}, \mathbb{1}[Z \in \mathcal{Z}_j(g)]\middle| X = x\right] \\
&\quad = \left(\mathbb{E}\left[\dot{Z}|Z \in \mathcal{Z}_j(g), X = x\right] - \mathbb{E}\left[\dot{Z}|Z \notin \mathcal{Z}_j(g), X = x\right]\right) \\
&\qquad \times \mathbb{P}[Z \in \mathcal{Z}_j(g)|X = x]\,\mathbb{P}[Z \notin \mathcal{Z}_j(g)|X = x], \qquad (29)
\end{aligned}$$

the sign of the correlation is given by the sign of

$$\mathbb{E}\left[\dot{Z}|Z \in \mathcal{Z}_j(g), X = x\right] - \mathbb{E}\left[\dot{Z}|Z \notin \mathcal{Z}_j(g), X = x\right],$$

whenever g, x, and j are such that $\mathbb{P}[Z \in \mathcal{Z}_j(g)|X = x] \in (0, 1)$.

Now suppose that the TSLS specification is monotonicity-correct for every $(z, \bar{z})$, and conditional on every x. Fix any g, x and j with $\mathbb{P}[G = g, X = x] > 0$ and $\mathbb{P}[Z \in \mathcal{Z}_j(g)|X = x] \in (0, 1)$, noting that (28) immediately implies that $\omega_j(g, x) = 0$ (and thus $\omega_j(g, x) \geq 0$) for all other combinations of g, x and j. Lemma 2 shows that $p(z_+, x) - p(z_-, x) > 0$ for every $z_+ \in \mathcal{Z}_j(g)$ and $z_- \notin \mathcal{Z}_j(g)$ with $\mathbb{P}[Z = z|X = x] \in (0, 1)$ for $z = z_-, z_+$. Since the TSLS specification is monotonicity-correct, these pairs must also satisfy

$$\dot{t}(z_+, x) - \dot{t}(z_-, x) = \gamma' i(z_+, x) - \gamma' i(z_-, x) \geq 0.$$

As a consequence,

$$\begin{aligned}\mathbb{E}\left[\dot{Z}|Z \in \mathcal{Z}_j(g), X = x\right] &= \mathbb{E}\left[\gamma' i(Z, x)|Z \in \mathcal{Z}_j(g), X = x\right] \\ &\geq \mathbb{E}\left[\gamma' i(Z, x)|Z \notin \mathcal{Z}_j(g), X = x\right] = \mathbb{E}\left[\dot{Z}|Z \notin \mathcal{Z}_j(g), X = x\right],\end{aligned}$$

and thus from (28)–(29), $\omega_j(g, x) \geq 0$, so that β_{tsls} is weakly causal, by Proposition 4.

Conversely, suppose that β_{tsls} is weakly causal, so that $\omega_j(g, x) \geq 0$ for all g, x, and $j \geq 1$ by Proposition 4. Note from (28) that $\omega_0(g, x) = 0$ as well, because $\mathbb{1}[Z \in \mathcal{Z}_j(g)] = 1$ deterministically when $j = 0$. Consider any value of x with $\mathbb{P}[X = x] > 0$. If $p(z, x)$ is constant as a function of z, then the TSLS specification is trivially monotonicity-correct for any pair $(z, \bar{z})$. So, suppose that there exists a pair (z_-, z_+) in the support of Z, conditional on $X = x$, such that $p(z_+, x) - p(z_-, x) > 0$.

Since $p(z_+, x) - p(z_-, x) > 0$, Lemma 2 implies that there exists a g and a j such that $\mathbb{P}[G = g|X = x] > 0$ with $z_+ \in \mathcal{Z}_j(g)$ and $z_- \notin \mathcal{Z}_j(g)$, so that $\mathbb{P}[Z \in \mathcal{Z}_j(g)|X = x] \in (0, 1)$. Because $\omega_j(g, x) \geq 0$ for all g, x, and j, (28)–(29) then imply that

$$\mathbb{E}\left[\dot{Z}|Z \in \mathcal{Z}_j(g), X = x\right] - \mathbb{E}\left[\dot{Z}|Z \notin \mathcal{Z}_j(g), X = x\right] \geq 0. \tag{30}$$

In order for (30) to be true, there must exist some $\bar{z}_+ \in \mathcal{Z}_j(g)$ and $\bar{z}_- \notin \mathcal{Z}_j(g)$—although not necessarily $(\bar{z}_-, \bar{z}_+) = (z_-, z_+)$—such that

$$\mathbb{E}\left[\dot{Z}|Z = \bar{z}_+, X = x\right] - \mathbb{E}\left[\dot{Z}|Z = \bar{z}_-, X = x\right] = \dot{t}(\bar{z}_+, x) - \dot{t}(\bar{z}_-, x) \geq 0.$$

Since $\bar{z}_+ \in \mathcal{Z}_j(g)$ and $\bar{z}_- \notin \mathcal{Z}_j(g)$, Lemma 2 implies that $p(\bar{z}_+, x) > p(\bar{z}_-, x)$. We conclude that the TSLS specification is monotonicity-correct for $(\bar{z}_-, \bar{z}_+)$ conditional on $X = x$. Because $x \in \mathcal{X}$ was arbitrary, this establishes the claim. *Q.E.D.*

Lemma 2. Suppose that Assumptions EX and WM are satisfied. Let (z_-, z_+) be two points in the support of Z, conditional on $X = x$. Let $\mathcal{Z}_j(x)$ denote the set of instrument values z for which individuals in choice group g would choose a treatment value t_j or larger. Then $p(z_+, x) - p(z_-, x) > 0$ if and only if there exists a choice group g with $\mathbb{P}[G = g|X = x] > 0$ and a treatment level t_j such that $z_+ \in \mathcal{Z}_j(g)$ and $z_- \notin \mathcal{Z}_j(g)$.

Proof of Lemma 2. Suppose that $p(z_+, x) > p(z_-, x)$. Assumption EX implies that

$$\mathbb{E}[T(z_+)|X = x] = p(z_+, x) > p(z_-, x) = \mathbb{E}[T(z_-)|X = x].$$

So given Assumption WM, it must be that

$$\mathbb{P}[T(z_+) \geq T(z_-)|X = x] = 1,$$

and hence that for some j,

$$\mathbb{P}[T(z_+) \geq t_j|X = x] > \mathbb{P}[T(z_-) \geq t_j|X = x] = \mathbb{P}[T(z_+) \geq t_j, T(z_-) \geq t_j|X = x].$$

It follows that

$$\begin{aligned} &\mathbb{P}[T(z_+) \geq t_j, T(z_-) < t_j|X = x] \\ &\quad = \mathbb{P}[T(z_+) \geq t_j|X = x] - \mathbb{P}[T(z_+) \geq t_j, T(z_-) \geq t_j|X = x] > 0. \end{aligned}$$

The definition of G implies that

$$\mathbb{P}[T(z_+) \geq t_j, T(z_-) < t_j|X = x] = \mathbb{P}[G \in \{g : z_+ \in \mathcal{Z}_j(g), z_- \notin \mathcal{Z}_j(g)\}|X = x],$$

so there must exist a g such that $\mathbb{P}[G = g|X = x] > 0$ with $z_+ \in \mathcal{Z}_j(g)$ and $z_- \notin \mathcal{Z}_j(g)$.

Conversely, suppose that such a g and treatment level t_j exist. For any $z_+ \in \mathcal{Z}_j(g)$ and $z_- \notin \mathcal{Z}_j(g)$ it follows that

$$\mathbb{P}[T(z_+) \geq t_j > T(z_-)|X = x] \geq \mathbb{P}[G = g|X = x] > 0,$$

noting in particular that this also implies $\mathbb{P}[T(z_+) > T(z_-)|X = x] > 0$. Assumption WM thus requires that

$$\mathbb{P}[T(z_+) \geq T(z_-)|X = x] = 1.$$

From Assumption EX it then follows that

$$p(z_+, x) = \mathbb{E}[T(z_+)|X = x] > \mathbb{E}[T(z_-)|X = x] = p(z_-, x).$$

Q.E.D.

Proof of Proposition 10. Assumption OSM implies that $p(z, x)$ is an increasing function of z for any x, so that $p(z_+, x) - p(z_-, x) \geq 0$ for any $z_+ \geq z_-$. The predicted value from the first stage regression is $\dot{t}(z_+, x) - \dot{t}(z_-, x) = (z_+ - z_-)\gamma$, so the specification will be monotonicity-correct for all pairs (z_-, z_+) if and only if $\gamma \geq 0$. From the Frisch-Waugh-Lovell Theorem, γ can be written as

$$\gamma = \frac{\mathbb{C}[T, \tilde{Z}]}{\mathbb{V}[\tilde{Z}]}.$$

The sign of γ is thus the same as the sign of

$$\begin{aligned} \mathbb{C}[T, \tilde{Z}] &= \mathbb{E}\left[\mathbb{E}\left[T|X, Z\right]\tilde{Z}\right] \\ &= \mathbb{E}\left[p(Z, X)(Z - \mathbb{E}[Z|X])\right] = \mathbb{E}\left[\mathbb{C}[p(Z, X), Z|X]\right], \end{aligned}$$

where in the first two equalities we used the assumption that covariates are rich with C containing a constant regressor. Because both $p(Z, X)$ and Z are increasing functions of Z, the covariance between $p(Z, X)$ and Z is positive conditional on X (e.g. Thorisson, 1995, Section 2), so $\mathbb{C}[T, \tilde{Z}]$ and thus γ are positive. The conclusion now follows from Proposition 9. *Q.E.D.*

Proof of Proposition 11. From Proposition 6, $\beta_{\text{tsls}} = \mathbb{E}[\tilde{Z}T]^{-1}\mathbb{E}[\tilde{Z}Y]$. Letting $\pi_g(x) \equiv \mathbb{P}[G = g|X = x]$, the numerator of β_{tsls} can be written as

$$\begin{aligned} \mathbb{E}[\tilde{Z}Y] &= \mathbb{E}\left[\sum_g \mathbb{E}\left[\tilde{Z}Y\big|X, G = g\right]\pi_g(X)\right] \\ &= \mathbb{E}\left[\sum_g \sum_{k=0}^{K} \mathbb{E}\left[\tilde{Z}Y\mathbb{1}[Z = \xi_k(X)]\big|X, G = g\right]\pi_g(X)\right] \\ &= \mathbb{E}\left[\sum_g \sum_{k=0}^{K} \mathbb{E}\left[\tilde{Z}Y\left(\tau(g, \xi_k(X))\right)\mathbb{1}[Z = \xi_k(X)]\big|X, G = g\right]\pi_g(X)\right], \end{aligned} \tag{31}$$

where $\tau(g, z)$ is the treatment level that choice group g chooses under instrument value

z. Assumption EX implies that

$$\begin{aligned}
&\mathbb{E}\left[\tilde{Z}Y\left(\tau(g,\xi_k(X))\right)\mathbb{1}[Z=\xi_k(X)]|X,G=g\right] \\
&\quad=\mathbb{E}\left[Y\left(\tau(g,\xi_k(X))\right)|X,G=g\right]\mathbb{E}\left[\tilde{Z}\mathbb{1}[Z=\xi_k(X)]\big|X,G=g\right] \\
&\quad=\mathbb{E}\left[Y|X,G=g,Z=\xi_k(X)\right]\mathbb{E}\left[\tilde{Z}\mathbb{1}[Z=\xi_k(X)]\big|X\right]. \qquad (32)
\end{aligned}$$

Inserting (32) into (31) yields

$$\begin{aligned}
\mathbb{E}[\tilde{Z}Y]&=\mathbb{E}\left[\sum_{k=0}^{K}\left(\sum_{g}\mathbb{E}\left[Y|X,G=g,Z=\xi_k(X)\right]\pi_g(X)\right)\mathbb{E}\left[\tilde{Z}\mathbb{1}[Z=\xi_k(X)]\big|X\right]\right] \\
&=\sum_{k=0}^{K}\mathbb{E}\left[\mathbb{E}[Y|X,Z=\xi_k(X)]\,\mathbb{E}\left[\tilde{Z}\mathbb{1}[Z=\xi_k(X)]\big|X\right]\right], \qquad (33)
\end{aligned}$$

where the second equality uses Assumption EX, which implies that $\pi_g(x)\equiv\mathbb{P}[G=g|X=x]=\mathbb{P}[G=g|X=x,Z=\xi_k(x)]$ for any k. Lemma 1 shows that (33) can also be written as

$$\mathbb{E}[\tilde{Z}Y]=\mathbb{E}\left[\mathbb{E}[Y|X,Z=\xi_0(X)]\,\mathbb{E}[\tilde{Z}|X]+\sum_{k=1}^{K}\Upsilon_k(X)\sum_{j=k}^{K}\mathbb{E}\left[\tilde{Z}\mathbb{1}[Z=\xi_k(X)]|X\right]\right],$$

If the TSLS specification is rich, so that $\mathbb{E}[\tilde{Z}|X=x]=0$ for each x, then this reduces to

$$\mathbb{E}[\tilde{Z}Y]=\mathbb{E}\left[\sum_{k=1}^{K}\Upsilon_k(X)\sum_{j=k}^{K}\mathbb{E}\left[\tilde{Z}\mathbb{1}[Z=\xi_k(X)]|X\right]\right].$$

Observe that with $\Xi_k(x)\equiv\{\xi_\ell(x):\ell\geq k\}$,

$$\begin{aligned}
\sum_{j=k}^{K}\mathbb{E}\left[\tilde{Z}\mathbb{1}[Z=\xi_k(X)]\big|X\right]&=\mathbb{E}\left[\tilde{Z}\mathbb{1}[Z\in\Xi_k(X)]\big|X\right] \\
&=\Big(\mathbb{E}[\tilde{Z}|Z\in\Xi_k(X),X]-\mathbb{E}[\tilde{Z}|Z\notin\Xi_k(X),X]\Big) \\
&\qquad\times\mathbb{P}[Z\in\Xi_k(X)|X]\,\mathbb{P}[Z\notin\Xi_k(X)|X] \\
&=\tilde{t}_k(X)\,\mathbb{P}[Z\in\Xi_k(X)|X]\,\mathbb{P}[Z\notin\Xi_k(X)|X],
\end{aligned}$$

so that

$$\mathbb{E}[\tilde{Z}Y] = \mathbb{E}\left[\sum_{k=1}^{K} \Upsilon_k(X)\tilde{t}_k(X)\,\mathbb{P}[Z \in \Xi_k(X)|X]\,\mathbb{P}[Z \notin \Xi_k(X)|X]\right].$$

Multiplying by $\mathbb{E}[\tilde{Z}T]^{-1} = \mathbb{E}[\tilde{Z}^2]^{-1}$ yields the stated expression for β_{rich}. *Q.E.D.*

B Finite sample performance of saturated specifications

B.1 Simulation design

We create a data generating process motivated by the empirical features of the data used in Card (1995), but with a binary treatment. We draw X uniformly from a discrete Halton sequence $\mathcal{X}$ on $[0, 1]$, the fineness of which we vary in different simulations. The instrument Z is then drawn conditionally on X according to

$$\mathbb{E}[Z|X = x] = \mathbb{P}[Z = 1|X = x] = 0.119 + 1.785x - 1.534x^2 + 0.597x^3. \tag{34}$$

The outcome and treatment are generating according to

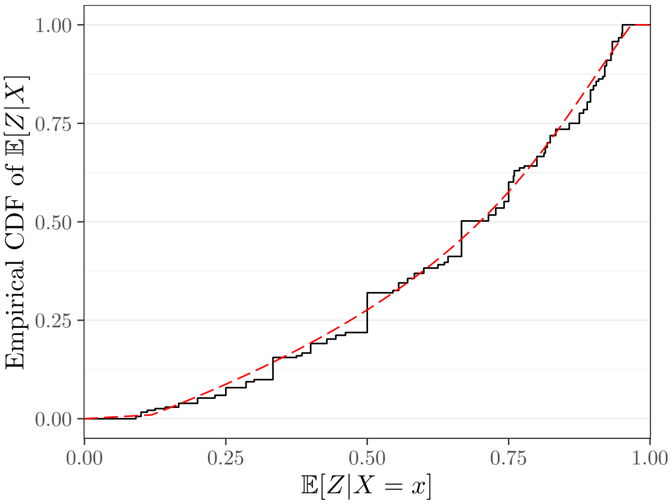
$$\begin{aligned} Y &= \log(129.7 + 1247.7X - 2149.0X^2 + 1515.7X^3) + 1.2T + U, \\ T &= \mathbb{1}[\Phi(V) \leq p(Z)], \end{aligned} \tag{35}$$

where (U, V) are standard multivariate normal with correlation .527, Φ is the standard normal distribution function (so that $\Phi(V)$ is uniformly distributed on $[0, 1]$), and the propensity score in our baseline is set to $p(0) = .22$ and $p(1) = .29$. The pair (U, V) is drawn independently of (X, Z).

This data generating process matches some key features of the data in Card (1995). The most important for our purposes is the estimated empirical distribution of $\mathbb{E}[Z|X]$, shown in black in Figure B.1, and plotted against the cumulative distribution function of $\mathbb{E}[Z|X = x]$, which is plotted in red. Our propensity score choices directly match the probability of college completion ($T = 1$, defined as 16 or more years of schooling) in the Card (1995) sample. The coefficient of 1.2 on the treatment indicator is the estimate obtained from the TSLS estimator of specification SS with a binary treatment. The correlation .527 between U and V is chosen to match the corresponding OLS estimator of specification SS.

Because the treatment enters additively in (35), treatment effects are constant both across observables and unobservables. As a consequence, the SS and SW specifications both have the same estimand, which is equal to the coefficient on T of 1.2. This allows

Figure B.1: Empirical distribution of college-presence instrument



Notes: This figure plots the distribution function of the conditional mean of the college-presence instrument used by Card (1995). The black line is the estimated empirical distribution of $\mathbb{E}[Z|X]$. The red line is $\mathbb{E}[Z|X = x]$ for the DGP in our simulations.

us to compare estimators based solely on their finite-sample distribution relative to a common target.

B.2 Baseline results

As a baseline, we take the number of covariates bins to be similar to the Card (1995) data (see Table 7) with $|\mathcal{X}| = 250$. Table B.1 reports results for four sample sizes, with the first row of each panel having 3,000 observations, similar to the Card (1995) data. All estimators of specification SW are highly biased, both in mean and in median, and are centered closer to the OLS estimator than they are to the true estimand of 1.2. This is expected for TSLS, but it is also true for IJIVE (Ackerberg and Devereux, 2009) and UJIVE (Kolesár, 2013), with the latter exhibiting a wildly noisy finite-sample distribution.[17] In contrast, the TSLS estimator of specification SS is less biased, although with a standard deviation that is considerably larger than the TSLS estimator of specification SW.

The other rows of Table B.1 show that the bias of the SW estimators decreases as n increases, consistent with the many instruments bias phenomenon. The jackknife

[17] The UJIVE estimator of specification SW is undefined when an (X, Z) bin has fewer than 2 observations, so we drop such bins when implementing it. We omit JIVE (Angrist et al., 1999) because it is expected to perform especially poorly with many covariates (Ackerberg and Devereux, 2009), which is what we found in simulations not reported here.

estimators improve much more quickly, with both showing less mean and median bias with n = 10,000 observations, and essentially none with n = 50,000. In contrast, the TSLS estimator remains severely biased even with n = 100,000 observations. All estimators become less variable as n increases, but the TSLS estimator of specification SS wins out in terms of root mean-squared error (RMSE) in all cases.

B.3 Varying instrument strength

Next, we vary the strength of the instrument by adjusting $p(1)$ while keeping $p(0) = .22$ fixed. We keep the number of covariate bins at the baselines of $|\mathcal{X}| = 250$.

Figure B.2 shows how the median estimate changes for each estimator as $p(1)$ increases, for four different sample sizes. Larger values of $p(1)$ and n lead to lower bias for all TSLS and JIVE estimators, as expected. Bias for the TSLS estimator of specification SS quickly disappears as the instrument gets stronger. In contrast, estimators of specification SW suffer from bias for much larger values of $p(1)$, suggesting a deleterious interaction between many and weak instruments (e.g. Chamberlain and Imbens, 2004; Chao and Swanson, 2005).

Table B.2 shows more details on the distributions of the five estimators for different instrument strengths with n = 3,000. The TSLS estimator of specification SW has much lower variance than the other TSLS and IJIVE estimators, but that's because overfitting in the first stage leads it to mimic OLS, and thus also be highly biased. The IJIVE and UJIVE estimators of specification SW are less biased, but also more variable. They end up being more variable than the TSLS estimator of specification SS, which consequently has the smallest RMSE in all cases except when the instrument is extremely weak ($p(1) - p(0) = .05$), in which case it is outperformed by OLS and the TSLS estimator of specification SW (which is essentially OLS).

B.4 Varying the number of covariate bins

Finally, we vary the number of covariate bins ($|\mathcal{X}|$) while keeping $p(0) = .22$, and $p(1) = .29$ fixed at their baseline values.

Figure B.3 shows how the median estimate changes for each estimator as $|\mathcal{X}|$ increases. Consistent with many instruments bias, the TSLS estimator of specification SW tends towards the OLS estimator as the number of groups increases, as do both IJIVE and UJIVE, but less quickly. In contrast, the TSLS estimator of specification SS is essentially median-unbiased even when there are an extremely large number of covariate bins.

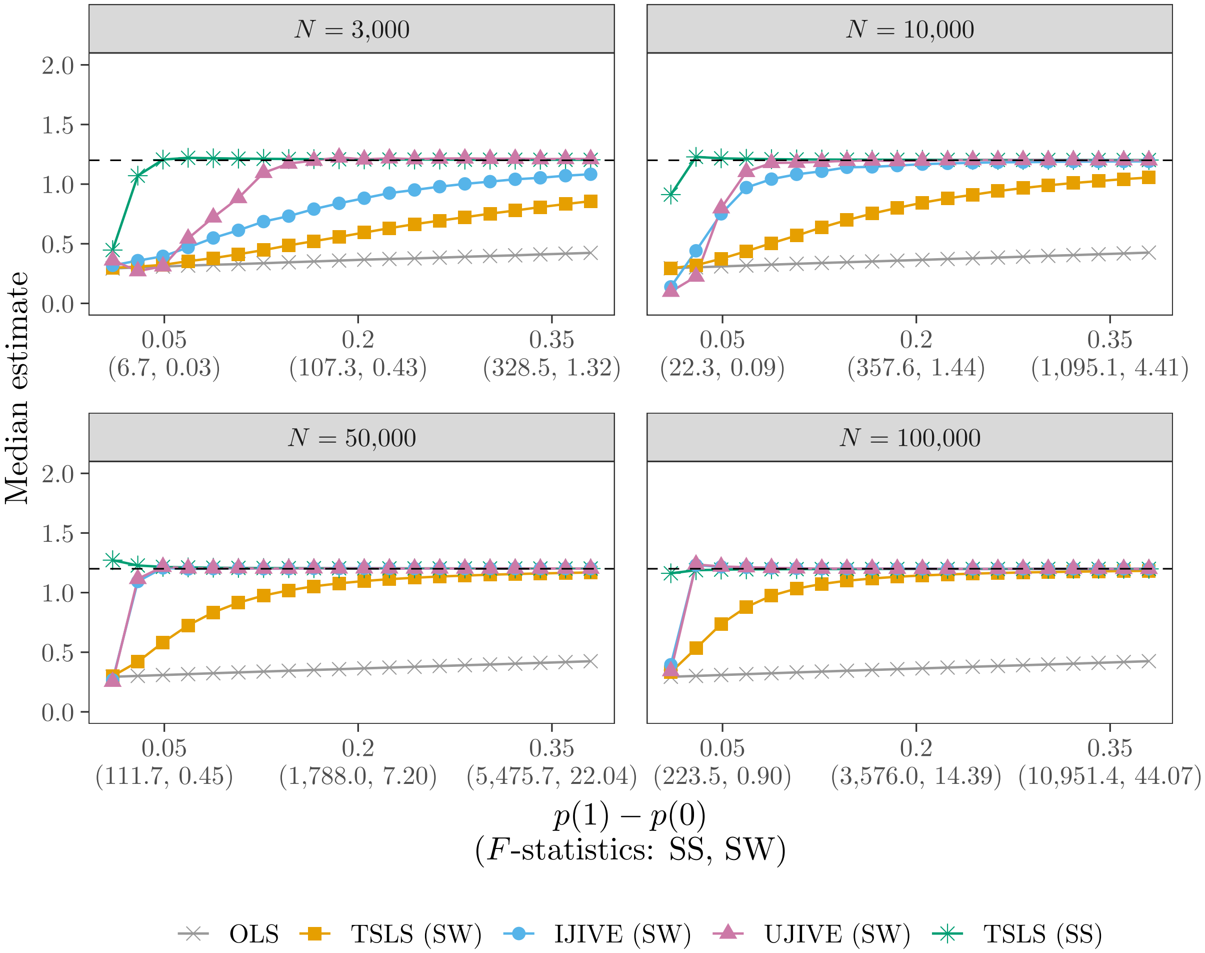
Table B.3 shows that the TSLS estimator of specification SW is both the most biased and least variable of the IV estimators. Its RMSE tends to be roughly comparable with that of the TSLS estimator of specification SS which has small bias, but larger variance. The IJIVE and UJIVE estimators of SW tend to be quite noisy, leading to RMSEs that are dominated by both the TSLS estimators of SW and SS.

Table B.1: Baseline with $|\mathcal{X}| = 250$, $p(0) = .22$, and $p(1) = .29$

N	Mean (SD)	RMSE	10%	25%	Median	75%	90%
	OLS						
3,000	0.317 (0.037)	0.884	0.265	0.289	0.318	0.344	0.367
10,000	0.317 (0.019)	0.883	0.291	0.304	0.317	0.331	0.342
50,000	0.317 (0.009)	0.883	0.306	0.311	0.317	0.324	0.329
100,000	0.317 (0.006)	0.883	0.309	0.312	0.317	0.322	0.326
	TSLS (SW)						
3,000	0.353 (0.134)	0.858	0.164	0.257	0.357	0.450	0.534
10,000	0.442 (0.123)	0.768	0.268	0.354	0.442	0.530	0.606
50,000	0.734 (0.091)	0.475	0.611	0.667	0.737	0.803	0.857
100,000	0.890 (0.082)	0.321	0.781	0.831	0.889	0.944	1.001
	IJIVE (SW)						
3,000	0.466 (0.578)	0.934	–0.276	0.118	0.479	0.822	1.196
10,000	1.146 (1.066)	1.067	0.093	0.528	0.966	1.541	2.431
50,000	1.205 (0.213)	0.213	0.920	1.048	1.198	1.347	1.501
100,000	1.209 (0.137)	0.137	1.028	1.108	1.213	1.298	1.401
	UJIVE (SW)						
3,000	0.688 (5.638)	5.661	–3.304	–0.773	0.562	1.976	4.746
10,000	1.472 (2.656)	2.670	–0.081	0.517	1.117	1.962	3.439
50,000	1.210 (0.215)	0.215	0.923	1.052	1.204	1.353	1.510
100,000	1.210 (0.137)	0.137	1.028	1.109	1.214	1.299	1.402
	TSLS (SS)						
3,000	1.300 (0.737)	0.744	0.377	0.797	1.220	1.733	2.264
10,000	1.224 (0.321)	0.322	0.815	0.982	1.212	1.435	1.655
50,000	1.206 (0.145)	0.145	1.010	1.099	1.211	1.308	1.398
100,000	1.202 (0.105)	0.105	1.066	1.123	1.195	1.278	1.353

Notes: Results are based on 1,000 repetitions. The true TSLS estimand is 1.2 for both specifications SW and SS. The mean, standard deviation, and root mean-squared error (RMSE) all exclude realizations smaller than the 1st percentile and larger than the 99th percentile.

Figure B.2: Varying the strength of the instrument with $|\mathcal{X}| = 250$



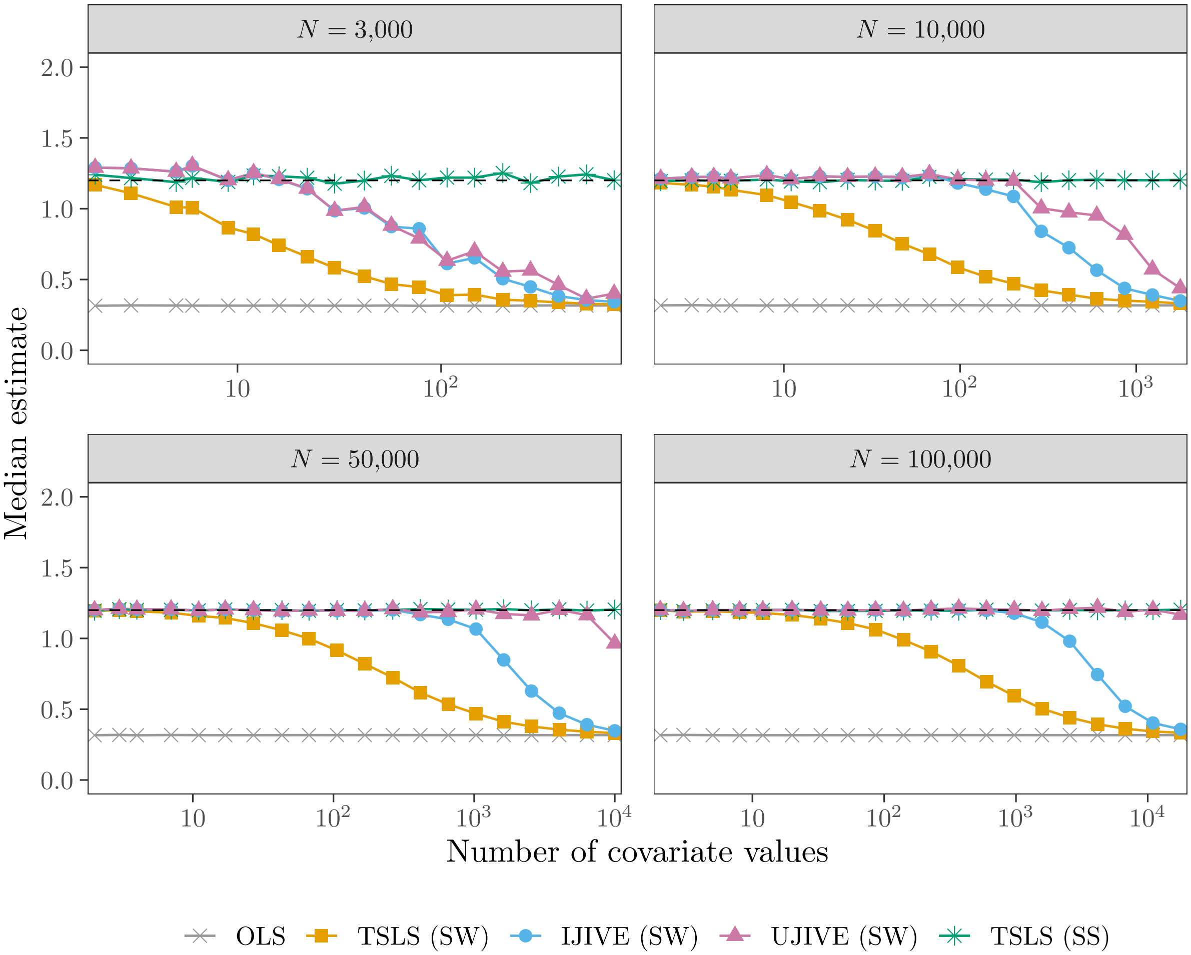
Notes: Each point represents the median estimate over 1,000 repetitions. The true estimand is 1.2 for both the SS and SW specifications. The reported F–statistics are computed from the population residuals and asymptotic variance matrix and indicated separately for specifications SS and SW.

Table B.2: Varying $p(1)$ with $n = 3{,}000$ and $|\mathcal{X}| = 250$

$p(1) - p(0)$	Mean (SD)	RMSE	10%	25%	Median	75%	90%
				OLS			
0.05	0.309 (0.038)	0.892	0.258	0.280	0.309	0.336	0.365
0.10	0.328 (0.036)	0.873	0.278	0.301	0.327	0.354	0.377
0.20	0.364 (0.035)	0.837	0.318	0.338	0.365	0.387	0.412
0.40	0.432 (0.033)	0.768	0.390	0.408	0.432	0.456	0.479
				TSLS (SW)			
0.05	0.326 (0.135)	0.884	0.140	0.228	0.324	0.425	0.501
0.10	0.401 (0.128)	0.809	0.224	0.309	0.401	0.490	0.578
0.20	0.586 (0.110)	0.623	0.436	0.506	0.585	0.660	0.742
0.40	0.882 (0.080)	0.328	0.771	0.829	0.880	0.940	0.991
				IJIVE (SW)			
0.05	0.393 (0.635)	1.026	–0.398	0.007	0.397	0.784	1.170
0.10	0.596 (0.469)	0.765	–0.001	0.271	0.590	0.900	1.226
0.20	0.886 (0.261)	0.408	0.547	0.689	0.872	1.064	1.248
0.40	1.094 (0.121)	0.161	0.932	1.008	1.095	1.183	1.255
				UJIVE (SW)			
0.05	0.120 (6.432)	6.522	–4.203	–1.064	0.316	1.912	4.439
0.10	0.568 (4.896)	4.937	–3.060	–0.360	0.773	1.885	4.350
0.20	1.297 (0.566)	0.574	0.654	0.888	1.204	1.577	2.115
0.40	1.212 (0.153)	0.154	1.002	1.100	1.212	1.326	1.426
				TSLS (SS)			
0.05	1.359 (1.265)	1.275	0.016	0.615	1.206	1.962	2.851
0.10	1.243 (0.469)	0.471	0.627	0.907	1.215	1.555	1.885
0.20	1.211 (0.222)	0.222	0.904	1.051	1.207	1.373	1.521
0.40	1.203 (0.109)	0.109	1.048	1.123	1.204	1.284	1.356

Notes: Same notes as for Table B.1.

Figure B.3: Varying $|\mathcal{X}|$ with $p(0) = .22, p(1) = .29$



Notes: Same notes as for Figure B.2. Note that the x-axis has a different scale in each facet.

Table B.3: Varying $|\mathcal{X}|$ with $n = 3{,}000$, $p(0) = .22$, and $p(1) = .29$

$\|\mathcal{X}\|$	Mean (SD)	RMSE	10%	25%	Median	75%	90%
	OLS						
200	0.316 (0.037)	0.885	0.266	0.289	0.316	0.342	0.365
100	0.316 (0.036)	0.885	0.267	0.291	0.315	0.341	0.366
50	0.316 (0.036)	0.885	0.268	0.290	0.315	0.341	0.364
10	0.315 (0.035)	0.885	0.266	0.289	0.315	0.341	0.364
	TSLS (SW)						
200	0.359 (0.137)	0.852	0.168	0.262	0.358	0.460	0.543
100	0.419 (0.185)	0.803	0.162	0.280	0.423	0.550	0.668
50	0.507 (0.258)	0.740	0.161	0.322	0.506	0.690	0.869
10	0.885 (0.445)	0.545	0.271	0.560	0.881	1.172	1.486
	IJIVE (SW)						
200	0.527 (0.784)	1.034	–0.366	0.082	0.517	0.971	1.447
100	1.074 (4.142)	4.144	–0.969	–0.041	0.747	1.542	3.272
50	1.134 (3.017)	3.017	–0.925	0.105	0.962	1.987	3.796
10	1.350 (1.766)	1.772	0.175	0.680	1.240	1.884	2.930
	UJIVE (SW)						
200	0.488 (6.297)	6.337	–3.596	–0.829	0.538	1.716	4.330
100	0.590 (5.198)	5.234	–2.569	–0.388	0.775	1.916	4.349
50	1.203 (4.385)	4.385	–1.414	0.052	0.980	2.052	4.114
10	1.359 (1.763)	1.771	0.175	0.680	1.242	1.888	2.945
	TSLS (SS)						
200	1.300 (0.707)	0.714	0.445	0.810	1.234	1.720	2.240
100	1.295 (0.681)	0.688	0.495	0.829	1.219	1.647	2.257
50	1.283 (0.673)	0.678	0.478	0.821	1.213	1.683	2.192
10	1.280 (0.622)	0.627	0.551	0.844	1.201	1.631	2.136

Notes: Same notes as for Table B.1.

C Specification SS can be monotonicity-incorrect even under Assumption SM

Suppose that $\mathcal{Z} = \{0, 1, 2\}$ and $\mathcal{X} = \{0, 1\}$, and take

$$C \equiv c(X) \equiv [1, X]' \quad \text{and} \quad I \equiv i(Z) \equiv [\mathbb{1}[Z = 1], \mathbb{1}[Z = 2]]' \equiv [Z_1, Z_2]'. \tag{36}$$

Then in the notation of Definition MC,

$$\dot{t}(2, x) - \dot{t}(1, x) = \gamma_2 - \gamma_1,$$

where $\gamma \equiv [\gamma_1, \gamma_2]'$ is the vector of population coefficients on I for the first stage regression. The claim is that even if $p(2, x) - p(1, x) > 0$ for both values of x, it is still possible to have $\gamma_2 - \gamma_1 < 0$, so that the TSLS first stage determined by (36) is not monotonicity-correct.

To see the intuition, let $V \equiv T - p(Z, X)$ be the difference between T and its conditional mean, then enumerate:

$$\begin{aligned} T &= p(0,0) + (p(0,1) - p(0,0))X + (p(1,0) - p(0,0))Z_1 + (p(2,0) - p(0,0))Z_2 \\ &\qquad (p(1,1) - p(0,1))Z_1X + (p(2,1) - p(0,1))Z_2X \\ &\equiv C'\lambda^\star + I'\gamma^\star + W'\zeta + V, \end{aligned}$$

where $W \equiv [Z_1X, Z_2X]'$ and the coefficient vectors collect the appropriate values of $p(z, x)$. Letting $\tilde{I} \equiv I - \mathbb{L}[I|C]$, $\tilde{T} \equiv T - \mathbb{L}[T|C]$, and $\tilde{W} \equiv W - \mathbb{L}[W|C]$, then applying the Frisch-Waugh-Lovell Theorem,

$$\gamma = \mathbb{E}[\tilde{I}\tilde{I}']^{-1}\,\mathbb{E}[\tilde{I}\tilde{T}] = \mathbb{E}[\tilde{I}\tilde{I}']^{-1}\,\mathbb{E}[\tilde{I}(\tilde{I}'\gamma^\star + \tilde{W}'\zeta + V)] = \gamma^\star + \underbrace{\mathbb{E}[\tilde{I}\tilde{I}']^{-1}\,\mathbb{E}[\tilde{I}\tilde{W}']\zeta}_{\text{omitted variables bias}}.$$

If the bias term is zero, then $\gamma = \gamma^\star$ and $\gamma_2 - \gamma_1 = p(2, 0) - p(1, 0) > 0$. However, the bias term is not zero in general.

As a numerical example, suppose that $\mathbb{P}[X = 1] = .5$, with

$$\mathbb{P}[Z = z|X = x] = \begin{cases} .5, & \text{if } z = 0 \\ .05 + .4x, & \text{if } z = 1 \\ .45 - .4x, & \text{if } z = 2 \end{cases}.$$

and set

$$p(z,0) = \begin{cases} 0, & \text{if } z = 0 \\ .085, & \text{if } z = 1 \\ .170, & \text{if } z = 2 \end{cases} \quad \text{and} \quad p(z,1) = \begin{cases} 0, & \text{if } z = 0 \\ .425, & \text{if } z = 1 \\ .510, & \text{if } z = 2 \end{cases}.$$

Then it can be shown through some tedious calculations that $\gamma = [.355, .24]'$, so that $\gamma_2 - \gamma_1 < 0$ even while $p(z,x)$ is strictly increasing in z for both values of x.

Intuitively, when $Z = 1$ it is overwhelmingly likely that $X = 1$, and when $Z = 2$, it is overwhelmingly likely that $X = 0$. So γ_1, the regression coefficient on Z_1, is mostly determined by variation in the $X = 1$ group, while γ_2, the regression coefficient on Z_2, is mostly driven by variation in the $X = 0$ group. Yet the change in the conditional mean of T from $Z = 0$ to $Z = 1$ conditional on $X = 1$ is much larger than the change from $Z = 0$ to $Z = 2$ conditional on $X = 2$. As a consequence, γ_1 ends up being larger than γ_2, violating monotonicity-correctness.

References

Abadie, A. (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113, 231–263. 3, 5, 13, 20, 21, 42

Ackerberg, D. A. and P. J. Devereux (2009): "Improved JIVE Estimators for Overidentified Linear Models with and without Heteroskedasticity," *The Review of Economics and Statistics*, 91, 351–362. 26, 56

Angrist, J. D. (1998): "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 66, 249–288. 16, 20, 21

——— (2001): "Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors," *Journal of Business & Economic Statistics*, 19, 2–28. 13

Angrist, J. D. and W. N. Evans (1998): "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *The American Economic Review*, 88, 450–477. 13, 27

Angrist, J. D., K. Graddy, and G. W. Imbens (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *The Review of Economic Studies*, 67, 499–527. 13

Angrist, J. D. and G. W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442. 2, 3, 12, 13, 21, 22, 25, 27, 29

——— (1999): "Comment on James J. Heckman, "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations"," *The Journal of Human Resources*, 34, 823. 2

Angrist, J. D., G. W. Imbens, and A. B. Krueger (1999): "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14, 57–67. 26, 56

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–455. 13, 14

Angrist, J. D. and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, 106, 979–1014. 4, 39, 40

——— (1999): "Chapter 23 Empirical Strategies in Labor Economics," Elsevier, vol. Volume 3, Part A, 1277–1366. 41

Angrist, J. D. and J.-S. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press. 2, 3, 4, 6, 10, 11, 12, 13, 16, 21, 25

Athey, S., J. Tibshirani, and S. Wager (2019): "Generalized Random Forests," *The Annals of Statistics*, 47, 1148–1178. 5, 42

Balke, A. and J. Pearl (1994): "Counterfactual Probabilities: Computational Methods, Bounds, and Applications," in *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI-94)*, ed. by R. Lopez de Mantras and D. Poole, 46–54. 13

——— (1997): "Bounds on Treatment Effects From Studies With Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171–1176. 13

Brinch, C. N., M. Mogstad, and M. Wiswall (2017): "Beyond LATE with a Discrete Instrument," *Journal of Political Economy*, 125, 985–1039. 42

Buckles, K. S. and D. M. Hungerman (2013): "Season of Birth and Later Outcomes: Old Questions, New Answers," *The Review of Economics and Statistics*, 95, 711–724. 40

Card, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed. by L. N. Christofides, K. E. Grant, and R. Swidinsky, Toronto: University of Toronto Press, 201–222. 4, 36, 37, 38, 39, 55, 56

Card, D., D. S. Lee, Z. Pei, and A. Weber (2015): "Inference on Causal Effects in a Generalized Regression Kink Design," *Econometrica*, 83, 2453–2483. 16

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101, 2754–81. 42

Chamberlain, G. and G. Imbens (2004): "Random Effects Estimators with many Instrumental Variables," *Econometrica*, 72, 295–306. 2, 11, 12, 57

Chao, J. C. and N. R. Swanson (2005): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73, 1673–1692. 57

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/Debiased Machine Learning for Treatment and Structural Parameters," *The Econometrics Journal*, 21, C1–C68. 5, 42

Deaton, A. (2010): "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature*, 48, 424–455. 2

Dube, O. and S. P. Harish (2020): "Queens," *Journal of Political Economy*, 128, 2579–2652. 4, 34, 35, 36, 37

Evdokimov, K. S. and M. Kolesár (2019): "Inference in Instrumental Variables Analysis with Heterogeneous Treatment Effects," *Working paper.* 18

Frölich, M. (2007): "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *Journal of Econometrics*, 139, 35–75. 5, 13, 25, 42

Gelbach, J. B. (2002): "Public Schooling for Young Children and Maternal Labor Supply," *American Economic Review*, 92, 307–322. 4, 30, 31, 32, 33, 35

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2021): "On Estimating Multiple Treatment Effects with Regression," *arXiv:2106.05024 [econ, stat]*. 16, 21

Goodman-Bacon, A. (2021): "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 225, 254–277. 2, 16

Heckman, J. (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *The Journal of Human Resources*, 32, 441–462. 2

Heckman, J., J. L. Tobias, and E. Vytlacil (2003): "Simple Estimators for Treatment Parameters in a Latent-Variable Framework," *Review of Economics and Statistics*, 85, 748–755. 42

Heckman, J. J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement.* 42

——— (1990): "Varieties of Selection Bias," *The American Economic Review*, 80, 313–318. 13

——— (2010): "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic Literature*, 48, 356–98. 25

Heckman, J. J. and R. Robb (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman and B. Singer, Cambridge University Press. 3, 21

Heckman, J. J., S. Urzua, and E. Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *Review of Economics and Statistics*, 88, 389–432. 25

Heckman, J. J. and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738. 25

Heckman, J. J. and E. J. Vytlacil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences of the United States of America*, 96, 4730–4734. 13, 42

Heiler, P. (2021): "Efficient Covariate Balancing for the Local Average Treatment Effect," *Journal of Business & Economic Statistics*, 1–14. 5, 42

Hirano, K., G. W. Imbens, D. B. Rubin, and X.-H. Zhou (2000): "Assessing the Effect of an Influenza Vaccine in an Encouragement Design," *Biostatistics*, 1, 69–88. 5, 13, 42

Hong, H. and D. Nekipelov (2010): "Semiparametric Efficiency in Nonlinear LATE Models," *Quantitative Economics*, 1, 279–304. 5, 42

Imbens, G. W. (2010): "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48, 399–423. 2

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. 2, 3, 5, 8, 13, 20, 22, 23, 42, 43

Imbens, G. W. and D. B. Rubin (1997): "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance," *The Annals of Statistics*, 25, 305–327. 5, 41

Kolesár, M. (2013): "Estimation in an Instrumental Variables Model with Treatment Effect Heterogeneity," *Working paper.* 3, 13, 18, 20, 25, 26, 56

Lam, D. A. and J. A. Miron (1987): "Seasonality of Births in Human Populations," Tech. Rep. 87-114, University of Michigan. 40

——— (1991): "Seasonality of Births in Human Populations," *Social Biology*, 38, 51–78. 40

LEE, D. S. (2008a): "Randomized Experiments from Non-Random Selection in U.S. House Elections," *Journal of Econometrics*, 142, 675–697. 16

LEE, J. (2008b): "Sibling Size and Investment in Children's Education: An Asian Instrument," *Journal of Population Economics*, 21, 855–875. 27

MANSKI, C. (1994): "The Selection Problem," in *Advances in Econometrics, Sixth World Congress*, vol. 1, 143–70. 13

MANSKI, C. F. (1989): "Anatomy of the Selection Problem," *The Journal of Human Resources*, 24, 343–360. 13

MANSKI, C. F. AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010. 42

MOGSTAD, M., A. SANTOS, AND A. TORGOVITSKY (2018): "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters," *Econometrica*, 86, 1589–1619. 42

OGBURN, E. L., A. ROTNITZKY, AND J. M. ROBINS (2015): "Doubly Robust Estimation of the Local Average Treatment Effect Curve," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77, 373–396. 5, 42

RAMSEY, J. B. (1969): "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis," *Journal of the Royal Statistical Society: Series B (Methodological)*, 31, 350–371. 13

ROBINS, J. M. AND S. GREENLAND (1996): "Identification of Causal Effects Using Instrumental Variables: Comment," *Journal of the American Statistical Association*, 91, 456–458. 2

ROSENZWEIG, M. R. AND K. I. WOLPIN (2000): "Natural "Natural Experiments" in Economics," *Journal of Economic Literature*, 38, 827–874. 27

SINGH, R. AND L. SUN (2022): "Double Robustness for Complier Parameters and a Semiparametric Test for Complier Characteristics," . 5, 42

SŁOCZYŃSKI, T. (2022): "When Should We (Not) Interpret Linear IV Estimands as LATE?" *arXiv:2011.06695 [econ, stat]*. 3, 13, 22, 24, 25, 26, 27, 42

SŁOCZYŃSKI, T., S. D. UYSAL, AND J. M. WOOLDRIDGE (2022): "Doubly Robust Estimation of Local Average Treatment Effects Using Inverse Probability Weighted Regression Adjustment," . 5, 42

SUN, B. AND Z. TAN (2021): "High-Dimensional Model-Assisted Inference for Local Average Treatment Effects With Instrumental Variables," *Journal of Business & Economic Statistics*, 1–13. 5, 42

SUN, L. AND S. ABRAHAM (2021): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 225, 175–199. 2, 16

SWANSON, S. A. AND M. A. HERNÁN (2014): "Think Globally, Act Globally: An Epidemiologist's Perspective on Instrumental Variable Estimation," *Statistical Science*, 29, 371–374. 2

Tan, Z. (2006): "Regression and Weighting Methods for Causal Inference Using Instrumental Variables," *Journal of the American Statistical Association*, 101, 1607–1618. 5, 13, 42

——— (2010): "Marginal and Nested Structural Models Using Instrumental Variables," *Journal of the American Statistical Association*, 105, 157–169. 5, 42

Thorisson, H. (1995): "Coupling Methods in Probability Theory," *Scandinavian Journal of Statistics*, 22, 159–182. 53

Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341. 13, 23

——— (2006): "Ordered Discrete-Choice Selection Models and Local Average Treatment Effect Assumptions: Equivalence, Nonequivalence, and Representation Results," *The Review of Economics and Statistics*, 88, 578–581. 13

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, MIT press. 13, 21

Yau, L. H. Y. and R. J. Little (2001): "Inference for the Complier-Average Causal Effect From Longitudinal Data Subject to Noncompliance and Missing Data, With Application to a Job Training Assessment for the Unemployed," *Journal of the American Statistical Association*, 96, 1232–1244. 5, 42